



Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Mortalidade Infantil no Distrito Federal: uma aplicação de Análise de Sobrevivência

Giovanna Valadares Borges

Orientador: Prof.^a Dr.^a Ana Maria Nogales Vasconcelos

Brasília

2018

Giovanna Valadares Borges

**Mortalidade Infantil no Distrito Federal:
uma aplicação de Análise de Sobrevivência**

Relatório apresentado à disciplina de Trabalho de Conclusão de Curso II de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Prof.^a Dr.^a Ana Maria Nogales Vasconcelos

Brasília

2018

Giovanna Valadares Borges

Mortalidade Infantil no Distrito Federal:
uma aplicação de Análise de Sobrevivência/ Giovanna Valadares Borges. – Brasília,
2018-

67 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof.^a Dr.^a Ana Maria Nogales Vasconcelos

Relatório Final – Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Trabalho de Conclusão de Curso de Graduação, 2018.

1. Introdução. 2. Objetivos. 3. Referencial Teórico. 4. Metodologia. 5. Resultados.
6. Conclusões.

Giovanna Valadares Borges

Mortalidade Infantil no Distrito Federal: uma aplicação de Análise de Sobrevivência

Relatório apresentado à disciplina de Trabalho de Conclusão de Curso II de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Trabalho aprovado. Brasília, 29 de Junho de 2018:

**Prof.^a Dr.^a Ana Maria Nogales
Vasconcelos**
Presidente da Banca

**Prof.^a Dr.^a Juliana Betini Fachini
Gomes**
Membro Interno - EST/UnB

Prof.^a MSc. Maria Teresa Leão Costa
Membro Interno - EST/UnB

Brasília
2018

*Dedico este trabalho ao meu avô, **Antônio Valadares** (in memorian).
A matemática não significaria o mesmo para mim sem você e eu não teria chegado até
aqui sem o seu incentivo e dedicação comigo. Saudade e Gratidão!*

Agradecimentos

Não é tarefa fácil agradecer a todos aqueles os quais gostaria aqui, mas me proponho a tentar.

Primeiramente, agradeço a Deus. Pelo dom da vida, pela família amorosa e cuidadosa em que me permitiu viver, por fazer de mim privilegiada com tantas oportunidades dadas e tantas pessoas incríveis que colocou em minha vida.

À minha família. Meus pais, Eduardo e Cláudia e minha avó, Clarice. Vocês sempre apoiaram incondicionalmente todos os meus sonhos e não há agradecimento suficiente por isso. Obrigada também por aguentarem todas as turbulências e os momentos de desespero. Vocês foram fundamentais para que eu não desistisse da minha caminhada nessa profissão. Amo vocês!

À minha orientadora, Ana Maria, que acreditou em mim em momentos que nem eu mesma acreditei. Você me inspira como profissional e como ser humano e sou muito grata pela orientação em todos os sentidos que me deu ao longo desses anos. Agradeço também pela parceria, que espero podermos perpetuar por muito tempo.

À professora Maria Teresa, por ter sido fonte de força e perseverança em momentos difíceis. Como já lhe disse inúmeras vezes, eu não sei o que seria de mim sem as diversas tardes em que fui à sua sala para aprender, conversar, desabafar, às vezes chorar, mas sempre ir embora sorrindo. Você me acolheu e, junto com a professora Ana Maria, colocou fé em mim e na profissional que estou me tornando. Serei eternamente grata por isso. Espero que a parceria (e as visitas) sejam frequentes daqui pra frente!

À professora Juliana, coordenadora do meu curso quando eu ingressei na UnB e hoje minha coorientadora. Você me auxiliou muito nos últimos semestres e eu sou muito grata pela compreensão, paciência, apoio e aprendizado que recebi de você.

Aos amigos da Católica, em especial Pedro Higor, Fernanda e Renan, que me apoiaram incondicionalmente quando resolvi encarar a Universidade antes do esperado. Obrigada pelos vários anos de amizade e por isso nunca ter mudado, independente da distância ou do tempo que se passe.

Aos muitos amigos que a UnB me trouxe, além dos amigos dos amigos. Em especial, aos amigos Lucas, Leylanne, Marina, Rodrigo, Pedro, Luiza, Ana Luiza, Matheus, Jady, Bárbara, entre tantos outros. Foram muitos dias (e muitas noites) juntos e, por isso, direciono a todos a minha gratidão pela amizade, pela força em momentos difíceis, pelas risadas nos finais de semana de estudos e por todas as histórias que vivemos.

Aos professores da Universidade de Brasília, não apenas do departamento de

Estatística, mas de todos que tive contato. Ao longo do tempo, percebi que a UnB é uma Universidade fora do padrão porque você também são. Encontrei profissionais incríveis, que me fizeram ficar cada vez mais apaixonada pelo que estava fazendo.

Aos colegas da Codeplan, onde estagiei no último ano, em especial à Rebeca, chefe compreensiva e amiga, que fez toda diferença no meu aprendizado como profissional e com seus conselhos de vida pessoal, e à Elisete, companheira de muitas risadas e conselhos acadêmicos. Vocês são profissionais incríveis e poder trabalhar ao lado de pessoas tão competentes e compromissadas me fez ver que esse sempre foi o tipo de profissional que eu quis ser. Espero ser como vocês quando crescer! Obrigada pela orientação, pelos almoços, as festinhas, as revisões de TCC (risos) e pela amizade que espero levar por muito anos.

Por fim, gostaria de agradecer à minha irmã, Raquel. Sempre brincamos que eu pedi uma irmãzinha para Deus e, sinceramente? Você foi o melhor presente que eu podia ter pedido a Ele. Obrigada por ser minha melhor amiga, companheira, conselheira, enfim, tudo o que você sabe que representa para mim. Acredito que estou trilhando o caminho que sonhei para mim e espero chegar longe. Mas espero mais ainda poder inspirar você a seguir seus sonhos e voar cada vez mais alto, sabendo que sempre estarei aqui para te apoiar em cada momento, assim como você sempre esteve presente por mim. Você é incrível e eu te amo muito!

"A morte não está mais perto do idoso do que do recém-nascido.

Nem a vida. "

(Khalil Gibran)

Lista de tabelas

Tabela 1 – Características da criança segundo presença de censura	40
Tabela 2 – Características gestacionais e de parto segundo presença de censura . .	40
Tabela 3 – Características maternas segundo presença de censura	41
Tabela 4 – Seleção de Covariáveis através da modelagem de Cox	47
Tabela 5 – Estimativas do modelo de Cox	51
Tabela 6 – Análise de Variância comparando os modelos APGAR e APGAR+IDANOMAL	51
Tabela 7 – Análise de Variância comparando os modelos IDANOMAL e AP- GAR+IDANOMAL	51
Tabela 8 – Coeficientes do modelo estimado completo	63
Tabela 9 – Coeficientes dos modelos estimados com 2 variáveis	63
Tabela 10 – Coeficientes dos modelos estimados com 3 variáveis	65
Tabela 11 – Coeficientes dos modelos estimados com 4 variáveis	67

Lista de ilustrações

Figura 1 – Tempo de sobrevivência do indivíduo até o óbito (em horas)	42
Figura 2 – Taxa de risco acumulado para ocorrência do óbito	43
Figura 3 – Tempo de sobrevivência segundo APGAR de 1º minuto.	44
Figura 4 – Tempo de sobrevivência segundo presença de anomalia ou malformação congénita.	44
Figura 5 – Tempo de sobrevivência segundo tipo de gravidez.	45
Figura 6 – Tempo de sobrevivência segundo número de filhos tidos mortos.	45
Figura 7 – Estimativa Cox-PH do modelo V1+V2 para $\hat{S}(t)$ segundo APGAR 1 .	48
Figura 8 – Estimativa Cox-PH do modelo V1+V2 para $\hat{S}(t)$ segundo APGAR 1 .	48
Figura 9 – Estimativa Cox-PH do modelo V1+V2+V4 para $\hat{S}(t)$ segundo APGAR 1	49
Figura 10 – Estimativa Cox-PH do modelo V1+V2+V4 para $\hat{S}(t)$ segundo IDANO- MAL	49
Figura 11 – Estimativa Cox-PH do modelo V1+V2+V4 para $\hat{S}(t)$ segundo GRAVIDEZ	50
Figura 12 – Resíduos de Cox-Snell para $\hat{S}(t)$	52
Figura 13 – Tempo de sobrevivência do indivíduo segundo sexo.	57
Figura 14 – Tempo de sobrevivência segundo peso ao nascer.	57
Figura 15 – Tempo de sobrevivência segundo APGAR de 5º minuto.	58
Figura 16 – Tempo de sobrevivência segundo tempo de gestação.	58
Figura 17 – Tempo de sobrevivência segundo nº de consultas pré-natais.	59
Figura 18 – Tempo de sobrevivência segundo tipo de parto.	59
Figura 19 – Tempo de sobrevivência segundo idade da mãe.	59
Figura 20 – Tempo de sobrevivência segundo raça/cor da mãe.	60
Figura 21 – Tempo de sobrevivência segundo escolaridade da mãe.	60
Figura 22 – Tempo de sobrevivência segundo estado civil da mãe.	60
Figura 23 – Tempo de sobrevivência segundo número de filhos tidos vivos.	61
Figura 24 – Tempo de sobrevivência segundo nº de partos tidos anteriormente (normais e/ou cesáreos).	61

Sumário

1	INTRODUÇÃO	17
2	OBJETIVOS	19
2.1	Objetivo Geral	19
2.2	Objetivos Específicos	19
2.3	Justificativa	19
3	REFERENCIAL TEÓRICO	21
3.1	Mortalidade Infantil: conceitos e realidade brasileira	21
3.1.1	Mortalidade infantil no Distrito Federal	22
3.1.2	Dados sobre mortalidade infantil	23
3.2	Análise de Sobrevivência	23
3.2.1	Conceitos básicos em Análise de Sobrevivência	24
3.2.1.1	Variável Aleatória	25
3.2.1.2	Função de Distribuição de Probabilidade	25
3.2.2	Estimação paramétrica das funções de Análise de Sobrevivência	27
3.2.2.1	Funções de Análise de Sobrevivência para variáveis aleatórias contínuas	27
3.2.3	Estimação não-paramétrica das funções de Análise de Sobrevivência	29
3.2.4	Modelo de regressão de Cox	31
4	METODOLOGIA	35
4.1	Bancos de Dados	35
4.2	Enfoque da pesquisa e preparação das bases	35
4.3	Técnica	37
5	RESULTADOS	39
5.1	Análise descritiva dos dados	39
5.1.1	Análise descritiva clássica	39
5.1.2	Análise descritiva da sobrevivência	42
5.2	Aplicação do modelo de regressão de Cox	46
6	CONCLUSÕES	53
	REFERÊNCIAS	55
	APÊNDICE A – GRÁFICOS DE SOBREVIVÊNCIA SEGUNDO CO-VARIÁVEIS	57

A.1	Tempos de sobrevivência segundo características da criança	57
A.2	Tempos de sobrevivência segundo características gestacionais e de parto	58
A.3	Tempos de sobrevivência segundo características maternas	59
	APÊNDICE B – MODELOS DE COX	63

1 Introdução

A mortalidade infantil compreende os óbitos de indivíduos menores de um ano de idade de certa população, dividida em: neonatal – subdividida entre neonatal precoce (óbitos ocorridos até 6º dia de vida) ou neonatal tardio (entre o 7º e o 27º dia) – e pós-neonatal (entre 28º e 364º dia) (SECRETARIA DE SAÚDE DO DISTRITO FEDERAL, 2015). A taxa de mortalidade infantil (TMI) é um dos principais indicadores de bem-estar social na atualidade (BARROS et al., 1985; SZWARCOWALD et al., 1997). Esse fenômeno sofreu uma forte queda em suas taxas entre a segunda metade do século passado e os dias atuais, levando ao surgimento de várias teorias para explicar as mudanças nos índices de mortalidade infantil no Brasil (COSTA et al., 2003).

No Distrito Federal, a TMI cai de 41 óbitos para cada mil nascidos vivos em 1980 (MONTEIRO; SCHMITZ, 2004) para apenas 10,6 em 2015, uma queda de 74,1%. Apesar disso, chama atenção que a maioria desses óbitos ocorreram no período neonatal e a principal causa são problemas na saúde da mãe ou complicações na gestação e/ou parto (SECRETARIA DE SAÚDE DO DISTRITO FEDERAL, 2015). Portanto, ainda com a diminuição da TMI nos últimos 30 anos, faz-se importante estudar a mortalidade infantil no DF. Sob o ponto de vista de uma Análise de Sobrevivência, determinar o tempo até a ocorrência do óbito e entender quais fatores podem determinar maior risco de um recém-nascido vir a falecer podem ser cruciais para o estabelecimento de práticas que venham a evitar óbitos infantis ao máximo no futuro.

O presente trabalho utiliza a Análise de Sobrevivência para avaliar a mortalidade infantil de residentes do Distrito Federal nos anos de 2013 a 2015. Os dados são do Sistema de Informações sobre Nascidos Vivos (SINASC) e o Sistema de Informações sobre Mortalidade (SIM), ambos fornecidos pelo Ministério da Saúde e seus respectivos gestores nos diferentes níveis - federal, estadual e municipal, através do DATASUS, plataforma online de disponibilização de dados socioeconômicos e demográficos.

Este trabalho está dividido em 6 seções. Esta primeira traz uma breve abordagem do trabalho. Na Seção 2, estão dispostos os objetivos gerais e específicos do trabalho e a justificativa para sua realização. A Seção 3, faz uma revisão da literatura sobre o problema de pesquisa e a técnica a ser utilizada na análise dos dados. A Seção 4 dispõe sobre a metodologia, com uma explanação sobre os dados utilizados, a variável de interesse, a aplicação da técnica e os softwares utilizados. Os resultados do trabalho são apresentados na Seção 5, com a estimação das curvas de sobrevivência e do risco, além do ajuste dos dados à um modelo probabilístico para análise de regressão. Por fim, a Seção 6 discute os resultados encontrados e as perspectivas futuras de continuidade da pesquisa.

2 Objetivos

2.1 Objetivo Geral

Utilizar a técnica de Análise de Sobrevivência para analisar a mortalidade infantil de residentes no Distrito Federal entre 2013 e 2015.

2.2 Objetivos Específicos

1. Contextualizar a mortalidade infantil no Brasil e no Distrito Federal;
2. Compreender a técnica de Análise de Sobrevivência;
3. Verificar a sobrevivência e as taxas de risco de óbito em crianças menores de 1 ano de idade, apontando o período em que o risco é mais expressivo;
4. Analisar, por meio do modelo de regressão de Cox, quais as variáveis que mais influenciam na ocorrência do óbito;
5. Avaliar se a modelagem de Cox tem boa aplicação para este conjunto de dados.

2.3 Justificativa

Utilizar a técnica de Análise de Sobrevivência aplicada à mortalidade infantil no Distrito Federal pode ser de grande contribuição para entender melhor o fenômeno. Com uso de modelo de regressão, será possível identificar características de maior influência nos óbitos, fundamentando futuras políticas públicas que possam auxiliar na diminuição das ocorrências de óbitos infantis.

3 Referencial Teórico

3.1 Mortalidade Infantil: conceitos e realidade brasileira

A qualidade de vida de uma sociedade é amplamente discutida na literatura, mas de difícil mensuração, ainda que avaliada por indicadores relacionados ao bem-estar da população (SZWARCWALD et al., 1997 apud THE WORLD BANK, 1995). No âmbito da saúde, a taxa de mortalidade infantil (TMI), ou coeficiente de mortalidade infantil, figura como um dos indicadores mais tradicionais quanto às condições de vida de uma população (BARROS et al., 1985; SZWARCWALD et al., 1997), “sendo considerada síntese da qualidade de vida e do nível de desenvolvimento de uma população. Assim, é possível estabelecer relações entre magnitude da TMI e as características do modelo de desenvolvimento” (CAMPOS; CARVALHO; BARCELLOS, 2000 apud PAIM; DIAS; ARAÚJO; BELIZZI, 1980, 1996).

Segundo Costa et al. (2003), até a década de 1970, a mortalidade infantil era vista como uma questão de subdesenvolvimento econômico, pois apresentava resultados inversos ao de indicadores econômicos, como o Produto Interno Bruto (PIB), a taxa de desemprego e o Índice de Gini. Essa relação foi proposta pela representatividade desses índices quanto às questões de renda, educação, saneamento e, conseqüentemente, saúde. Assim, um país com alto rendimento e solução para o desemprego e desigualdade social teria, em tese, um baixo índice de mortalidade infantil.

Tal argumentação perdeu força no final da mesma década, já que países em desenvolvimento tiveram quedas persistente nos seus índices de mortalidade infantil mesmo vivendo crises econômicas que envolveram redução do PIB, desvalorização dos salários e aumento do desemprego. Essa ruptura com as ideias anteriormente propagadas estimulou uma nova onda de estudos científicos que trouxessem explicações sobre o declínio dos índices de mortalidade infantil, mesmo em países em desenvolvimento.

Ainda assim, a relação entre desenvolvimento socioeconômico e a mortalidade infantil é uma das bases da teoria da Transição Demográfica, pois essa descreve que as mudanças na natalidade e mortalidade envolvem todo o processo de desenvolvimento socioeconômico, com a modernização das sociedades. Essa transição foi verificada em países europeus, com conseqüentes mudanças nos ritmos de crescimento populacional (VASCONCELOS; GOMES, 2012).

Mosley e Chen (1984) propõem uma abordagem analítica de estudo dos determinantes da mortalidade infantil nos países em desenvolvimento. Os autores mostram a importância em considerar tanto os fatores sociais quanto médicos na formulação de

modelo explicativo para o fenômeno. De um lado, os estudos sociais focam na relação entre nível socioeconômico e padrões de mortalidade, produzindo inferências causais sobre os determinantes de mortalidade. Renda e nível de instrução da mãe entram como um exemplo clássico de fatores comumente relacionados ao óbito infantil. Já o outro foca no processo biológico, correlacionando a mortalidade com quadros clínicos como infecções ou malnutrição.

Ainda que ambas contribuam para a compreensão do fenômeno ao longo do tempo, as diferentes abordagens e metodologias de trabalho distanciam as interpretações e contêm o desenvolvimento de ferramentas mais efetivas para análise da mortalidade. Desta forma, combinar os determinantes médicos e socioeconômicos é um fator de grande contribuição para o desenvolvimento de um modelo analítico da mortalidade.

Os autores apresentam 5 categorias de variáveis que combinam questões médicas e sociais para determinação do modelo: fatores maternos, contaminação do ambiente, deficiência nutricional, danos acidentais ou intencionais à saúde e controle de doenças. De fato, Monteiro e Schmitz (2004) mostram que, entre 1990 e 2000, as principais causas de mortalidade neonatal (óbitos ocorridos até 27º dia de vida) são problemas congênitos da criança, complicações na saúde da mãe ou durante a gestação e parto. Diferente de outros países, o Brasil apresenta um nível muito alto de óbitos neonatais devido à problemas ao longo da gestação ou parto. Nos óbitos pós-neonatal, vê-se alta ocorrência de fatores ambientais como causa, tais como infecções respiratórias e intestinais e deficiências nutricionais. Doenças infecto-parasitárias, tais como diarreias e septicemias, também se apresentam como causa desses óbitos.

3.1.1 Mortalidade infantil no Distrito Federal

Seguindo a tendência a nível federal, a mortalidade infantil no Distrito Federal tem padrão de queda gradativa, com a taxa de mortalidade indo de 41 óbitos por mil nascidos vivos em 1980, para 26,3 em 1990 e 14,4 em 2000 (MONTEIRO; SCHMITZ, 2004). Já em 2015, a Secretaria de Saúde do Distrito Federal (2015) mostra que esse índice cai para 10,6, o menor nível já registrado no DF.

Nesse mesmo relatório vemos que, entre 2013 e 2015, mais de 70% dos óbitos infantis ocorrem no período neonatal e cerca de 30% do total de óbitos foram de crianças com peso entre 500g e 1kg. Além disso, 62,4% dos óbitos infantis foram por problemas perinatais, como fatores de saúde materna e complicações na gravidez e parto, e 25,9% por malformação congênita, um resultado parecido com de Monteiro e Schmitz (2004) e concordante com a modelagem de Mosley e Chen (1984). Isso mostra que a melhoria no acesso e à qualidade de assistência pré e pós natal permanece necessária tanto a nível federal quanto em nível distrital.

3.1.2 Dados sobre mortalidade infantil

Com relação à obtenção de dados e a mensuração da mortalidade infantil, o coeficiente é calculado com base em dados do Registro Civil e de pesquisas demográficas, como o Censo. Outras fontes de dados de saúde pública sobre natalidade e mortalidade são o SINASC e o SIM.

É importante ressaltar que ambas as fontes passaram por diversas reestruturações ao longo do tempo, o que afetou fortemente a qualidade dos dados. No entanto, nos últimos anos, houve uma melhora na obtenção e na estruturação desses dados, permitindo que se tornassem ferramentas úteis e mais fidedignas para avaliação da saúde pública no Brasil.

3.2 Análise de Sobrevivência

A Análise de Sobrevivência, ou Confiabilidade, é um ramo da Estatística que tem como variável de interesse o tempo até a ocorrência de um determinado evento. Esse evento é conhecido como falha. Em pesquisas médicas, por exemplo, o interesse pode ser o tempo até o óbito do paciente ou até a cura de uma determinada doença, análises estas que generalizaram o uso do termo “sobrevivência”.

Já em pesquisas industriais, onde o termo “confiabilidade” é mais comum, o interesse pode ser o tempo até a falha de uma máquina, de um produto ou tempo decorrido até a necessidade de reparo. Esse tipo de análise pode ser aplicado em diversas outras áreas, como ciências atuariais, economia, educação, etc.

A variável tempo até o evento é estritamente positiva e pode ser medida em diversas unidades, geralmente como uma variável contínua. No entanto, trabalhar com acompanhamento temporal pode ser complicado, uma vez que muitas observações podem ser consideradas perdidas por não ser possível continuar captando as informações do indivíduo ou produto em estudo. Por exemplo, um paciente participante de uma pesquisa para estudar as características de uma determinada patologia pode deixar de frequentar as avaliações periódicas com os médicos, antes que a cura de sua doença seja observada. Ou em um teste para verificar o tempo de falha até um componente, nem todos os componentes falham até o final do estudo. A questão é que as informações captadas anteriormente não terão continuidade, ou seja, perderemos a informação pela não-observação do fenômeno.

Em Análise de Sobrevivência, tem-se a possibilidade de trabalhar com esses dados, que serão tratados como censurados. A censura é a observação parcial de uma resposta, importante e útil para esse tipo de análise. Como resalta (COLOSIMO; GIOLO, 2006), todas as observações, ainda que censuradas, devem ser consideradas nas estatísticas de um estudo que utilize Análise de Sobrevivência pois, ainda que incompletas, as censuras fornecem informações sobre o tempo de sobrevivência. Além disso, ao omitir os dados

censurados, as estatísticas do estudo podem gerar conclusões viesadas.

Assim, deve-se acrescentar uma variável indicadora de censura γ_i ao conjunto de dados. É atribuído o valor 1 às observações cujo o tempo de sobrevivência foi observado e valor 0 àquelas que foram censuradas, tal como:

$$\delta_i = \begin{cases} 0 & \text{se } i - \text{esimo tempo foi censurado} \\ 1 & \text{se } i - \text{esimo tempo foi observado} \end{cases}, i = 1, 2, \dots$$

As censuras podem ser classificadas em censura à direita, à esquerda e intervalar. Na censura à direita, tem-se que o tempo de interesse está à direita do tempo registrado. Já na censura à esquerda, o tempo de interesse está à esquerda do tempo registrado. Por fim, a censura intervalar ocorre quando são feitas visitas regulares aos indivíduos em análise e o tempo de interesse ocorreu no período entre as visitas.

Divide-se também a censura à direita em censura do tipo I, do tipo II e aleatória. Na censura do tipo I, o tempo de falha de alguns indivíduos é menor do que o tempo pré-determinado do estudo. Assim, a observação é censurada por não se saber quando ou se a falha irá ocorrer por não-observação. Nesse caso, o tempo é determinado e o número de censuras captadas pela pesquisa é aleatório.

Na censura do tipo II, define-se um número k de censuras a serem captadas (sendo $k \leq n$, onde n é o tamanho da amostra) pelo estudo. Desta forma, a pesquisa irá ocorrer até que o número pré-definido de falhas ocorram, fazendo do tempo uma variável aleatória.

Por fim, a censura aleatória ocorre de maneira natural, sem intervenções do pesquisador. É a forma generalizada e acaba englobando as outras duas, pois a censura pode ocorrer quando os indivíduos não falharem até o fim da pesquisa ou por motivos alheios e imprevisíveis à pesquisa, como morte, mudança de cidade, perda de contato com o indivíduo, entre outros.

É interessante considerar também a presença de variáveis explicativas, ou covariáveis, aquelas observadas além do tempo de sobrevivência e da indicação de censura. Essas variáveis podem ter relação e até influenciar o tempo de sobrevivência do indivíduo que poderão ser estudadas através de modelos de regressão como variáveis explicativas do tempo até a ocorrência do evento.

3.2.1 Conceitos básicos em Análise de Sobrevivência

Afim de possibilitar a descrição do tempo matematicamente, é necessário defini-lo em função. Em geral, descrevemos o tempo de falha T como uma variável aleatória não-negativa contínua. Mas se considerarmos $T = [Y]$, sendo $[Y]$ a “parte inteira de Y ” (maior inteiro menor ou igual a Y), podemos definir o tempo como uma variável discreta. Com isso, é necessário adaptar o método de análise para comportar funções discretas sem

acarretar em perdas ou vieses. Esse desafio tem sido trabalhado na literatura, em trabalhos como de Nakano e Carrasco (2006).

Desta forma, serão descritas as funções utilizadas em Análise de Sobrevida na forma contínua, uma vez que o foco do trabalho será analisar o tempo até o óbito em horas. As definições dadas a seguir são fundamentais para apresentação do modelo de Análise de Sobrevida e foram retiradas do trabalho de Magalhães (2006).

3.2.1.1 Variável Aleatória

Considere um determinado fenômeno aleatório que se deseja estudar. Antes de sua realização, exceto em raros casos, não sabemos seu resultado. Mas podemos convenientemente estabelecer seu espaço de probabilidade e avaliar a probabilidade de qualquer evento de interesse da σ -álgebra, permitindo a atribuição de probabilidades às funções desses eventos.

Assim, seja $(\Omega, \mathcal{F}, \mathcal{P})$ um espaço de probabilidade. É considerada uma variável aleatória qualquer função $T : \Omega \rightarrow \mathbb{R}$ tal que:

$$T^{-1}(I) = \{w \in \Omega : T(w) \in I\} \in \mathcal{F}, \forall \text{ intervalo } I \subset \mathbb{R}$$

Ou seja, T é tal que sua imagem inversa do intervalo $I \subset \mathbb{R}$ pertencem à σ -álgebra \mathcal{F} . Desta forma, dizemos que uma variável aleatória é uma função do espaço amostral Ω em \mathbb{R} a qual tem-se a possibilidade de calcular a probabilidade de ocorrência de seus valores.

3.2.1.2 Função de Distribuição de Probabilidade

Conhecer a função de distribuição de uma dada variável aleatória permite obter qualquer informação sobre a variável, sejam quaisquer os valores assumidos num subconjunto em \mathbb{R} , uma vez que essa é definida por toda a reta. Assim, seja T uma variável aleatória em $(\Omega, \mathcal{F}, \mathcal{P})$, sua função de distribuição é dada por:

$$F_T(t) = P(T \in [-\infty, t]) = P(T \leq t)$$

Ou seja, a distribuição é encontrada através da probabilidade da variável T assumir valores dentro do intervalo definido entre $-\infty$ e um determinado valor t . Essa função obedece às seguintes propriedades:

1.

$$\lim_{t \rightarrow -\infty} F_T(t) = 0 \text{ e } \lim_{t \rightarrow \infty} F_T(t) = 1;$$

2. F_T é contínua à direita;

3. F_T é não decrescente, isto é, $F_T(t) \leq F_T(y)$ sempre que $t \leq y, \forall t, y \in \mathbb{R}$.

Antes de definir uma função, é importante identificar a classificação da variável aleatória de acordo com os valores que ela pode assumir, uma vez que o cálculo de probabilidade da mesma está relacionado à essa classificação. Uma variável que assume apenas valores enumeráveis, seja o conjunto desses valores finito ou infinito, é dita discreta. Um exemplo deste tipo de variável seria o número de evasões em um determinado colégio. Nesse caso, não existe “meia” evasão, existem apenas t_i valores pontuais, onde i assume valores $1, 2, 3, \dots$

Para esse tipo de variável, temos que a função de distribuição dos valores é descontínua (pontual), com formato de escada, em que cada um dos pontos de descontinuidade são aqueles assumidos pela variável. Assim, a função de probabilidade será uma função cuja probabilidade é atribuída a cada um dos valores que a variável pode assumir. Portanto, seja T uma variável com valores t_i , $i = 1, 2, \dots$, temos:

$$p(t_i) = P(T = t_i) = P(w \in \Omega : T(w) = t_i)$$

Tratando-se de uma probabilidade, duas condições devem ser respeitadas:

1. $0 \leq p(t_i) \leq 1, \forall i$;
2. $\sum_i p(t_i) = 1$, com a soma percorrendo todos os possíveis valores para i .

Sendo discreta, a variável terá sua função de distribuição equivalente à função de probabilidade e, para qualquer $t \in \mathbb{R}$, podemos escrever:

$$F_T(t) = \sum_{i \in A_t} p(t_i), \text{ com } A_t = i : t_i \leq t$$

Uma variável será considerada contínua caso exista uma função não negativa f de modo que para todo $t \in \mathbb{R}$ seja válido:

$$F_T(t) = \int_{-\infty}^t f(w) dw$$

Neste caso, passamos a considerar um intervalo de valores a serem assumidos pela variável. Essa função também é conhecida como função de distribuição acumulada. Um exemplo desse tipo de variável é o tempo até o gasto total de uma lâmpada. Em geral, contabilizamos em horas, minutos, segundos, mas há um conjunto não enumerável de tempo entre 0 e 1 minuto, 1 e 2 horas, etc. Assim, tomamos um intervalo de tempo como formar de contar e avaliar sua duração.

Diferente do caso discreto, essa função de probabilidade não equivale à função de distribuição e, nesse caso, a função F é a função de distribuição, a função f é chamada função densidade. Ainda assim, assume propriedades semelhantes à função de probabilidade discreta:

1. $f(t) \geq 0, \forall t \in \mathbb{R};$
2. $\int_{-\infty}^{\infty} f(w) dw = 1.$

É interessante notar que a função densidade é importante para a caracterização de uma variável contínua e que, ainda que não seja semelhante à função de distribuição, ambas guardam relação matemática. Assim, da mesma forma que obtemos $F_T(t)$ a partir da integração da função densidade, podemos obter $f_T(t)$ derivando $F_T(t)$, tal que:

$$f_T(t) = \frac{d}{dt} F_T(t)$$

3.2.2 Estimativa paramétrica das funções de Análise de Sobrevida

Em um modelo de Análise de Sobrevida, temos que a variável aleatória estudada é o tempo até a ocorrência de um evento de interesse, também chamado de falha. Esse fenômeno não é necessariamente um fracasso. Um exemplo seria a análise do tempo até a cura de uma determinada doença, tendo assim a cura como o evento de interesse e o tempo até a sua ocorrência a variável que desejamos estudar.

Assim, sendo a distribuição de probabilidade desse evento conhecida, bem como apresentado anteriormente, poderemos especificar o comportamento da variável aleatória através das funções de sobrevivência e de risco (também chamada de taxa de falha). Em geral, o tempo será considerado uma variável aleatória contínua e, assim como há uma relação matemática entre a função de distribuição e a função densidade desse tipo de variável, a função de sobrevivência e a função de risco também estarão relacionadas entre si, de tal forma que, conhecendo uma, podemos usar o processo de integração ou derivada para obter a outra. Há também a possibilidade de estudar variáveis discretas por meio dessa técnica, mas a relação entre as funções permanece, obtida através de outras relações matemáticas.

3.2.2.1 Funções de Análise de Sobrevida para variáveis aleatórias contínuas

A partir de Colosimo e Giolo (2006) e Nakano (2017), escrevemos a função de sobrevivência como a probabilidade de um indivíduo não falhar até um determinado tempo t , definida por:

$$S(t) = P[T > t] = \int_t^{\infty} f(t) dt, \quad t \geq 0$$

Essa função é não-crescente, absolutamente contínua e deve ter que

$$\lim_{t \rightarrow -\infty} S(t) = 0 \text{ e } \lim_{t \rightarrow \infty} S(t) = 1;$$

Já a função de risco (ou taxa de falha) irá representar o risco instantâneo do indivíduo vir a sofrer o evento de interesse, em um determinado tempo t . No caso contínuo,

essa função é definida por: $h(t) =$

$$\lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, t \geq 0$$

Ou seja, trata-se da razão do limite da probabilidade de um indivíduo experimentar o evento de interesse no intervalo $[t, t + \Delta t]$ dado que ele não tenha sofrido o evento até o tempo t . Assim, a função de risco é estritamente positiva e, diferente da função de sobrevivência, não é limitada superiormente.

Essa função guarda relação matemática com a função densidade e a função de sobrevivência, nos permitindo encontrá-la mais facilmente a partir da seguinte equação:

$$h(t) = \frac{f(t)}{S(t)}$$

$h(t)$, portanto, descreve como o risco se modifica ao longo do tempo e Colosimo e Giolo (2006) discutem como alguns autores consideram a função de risco mais informativa do que a função de sobrevivência, pois ainda que funções de sobrevivência de diferentes distribuições tenham comportamentos parecidos, a função de risco poderá ser totalmente diferente.

A função de risco pode apresentar graficamente comportamento constante, monótono (crescente ou decrescente) e não monótono (unimodais, multimodais ou em formato de "u") ao longo do tempo. Entende-se que quando a função tem comportamento constante, a taxa de falha não se altera a medida que t cresce. No caso monótono, se ela é crescente, a chance de falhar aumenta com o tempo. Se é decrescente, essa chance diminui. Nos casos não monótonos, o comportamento da taxa de falha tem diferentes variações ao longo do tempo. Se é unimodal, entende-se que é crescente até certo ponto e depois torna-se decrescente. Se tem formato de "u", ou de banheira, tem comportamento contrário ao unimodal. Ou seja, é decrescente nos tempos iniciais e, em certo ponto, passa a ser crescente. Por fim, se é multimodal, tem momentos de crescimento seguidos de decrescimento ao longo do tempo.

No entanto, em procedimentos de estimação não-paramétricos e de seleção de modelos, torna-se mais útil utilizar a Função de Risco Acumulada (ou Taxa de Falha Acumulada), denotada por $H(T)$. Ainda que não possua interpretação direta, podemos dizer que ela fornece o risco acumulado do indivíduo sofrer o evento de interesse até o tempo t , representado por:

$$H(t) = \int_0^t h(u) du, t \geq 0$$

Essa função também guarda relação matemática com a função de sobrevivência, dada por:

$$S(t) = \exp [-H(t)]$$

3.2.3 Estimação não-paramétrica das funções de Análise de Sobrevivência

Em muitas situações, a distribuição de probabilidade do evento não é conhecida. No entanto, como ainda desejamos estudar esses casos, existem métodos que nos auxiliam na obtenção de resultados aproximados em um modelo de Sobrevivência, chamada estimação não-paramétrica. Por não-paramétrica, quer-se justamente dizer que não há uma distribuição de probabilidade conhecida para aquele evento, portanto não se conhecem os parâmetros que determinam a distribuição. A estimação é feita com base nas observações de uma amostra conhecida, obtendo-se as funções de Sobrevivência apresentadas em seção anterior.

Quando trabalhamos com estimação em Sobrevivência, devemos considerar dois cenários: se estamos trabalhando com um conjunto de dados sem nenhuma censura ou se existem observações censuradas. Caso não existam, a função de sobrevivência pode ser estimada simplesmente por:

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações com tempo de sobrevivência } > t}{\text{total de observações}}, \forall t \geq 0$$

No entanto, na presença de censura, não podemos utilizar tal função pois esta ignora qualquer informação sobre indivíduos com tempo de sobrevivência maior que o considerado, mas que tenham sofrido censura antes desse momento. É por isso que, em 1985, Kaplan e Meier desenvolvem um método de estimação não-paramétrica para sobrevivência quando há presença de censura na amostra estudada, conhecido como estimador de Kaplan-Meier (BASTOS; ROCHA, 2007).

O estimador de Kaplan-Meier é uma adaptação da estimação anterior para incluir as censuras. Assim, considerando que se um indivíduo sobreviveu até o tempo j , ele necessariamente sobreviveu no tempo $j - 1$. Desta forma, a probabilidade de sobrevivência pode ser escrita como:

$$\begin{aligned} S(t_{(j)}) &= P(T > t_{(j)}) \\ &= P(T > t_{(j-1)}, T > t_{(j)}) \\ &= P(T > t_{(j-1)})P(T > t_{(j)}|T > t_{(j-1)}) \\ &= S(t_{(j-1)})[1 - P(T > t_{(j)}|T > t_{(j-1)})] \\ &= S(t_{(j-1)})[1 - P(t_{(j-1)} < T < t_{(j)}|T > t_{(j-1)})] \\ &= S(t_{(j-1)})[1 - P(t_{(j-1)} < T < t_{(j)}|T > t_{(j)})] \end{aligned}$$

Assim, temos que a função de sobrevivência será:

$$S(t) = S(t_{(0)}) \prod_{j:t_{(j)} \leq t} [1 - P(t_{(j-1)} < T < t_{(j)} | T > t_{(j)})]$$

Sendo que a probabilidade de sobreviver no tempo 0 é 1, temos que $S(t_0) = 1$, a estimativa de Kaplan-Meier irá estimar a probabilidade $\hat{q}_j = P(t_{(j-1)} < T < t_{(j)} | T > t_{(j)})$ de um indivíduo vir a experimentar o evento em um intervalo $(t_{(j-1)}, t_{(j)})$ estando sob risco em $T_{(j)}$ (NAKANO, 2017). Assim, podemos dizer que:

$$\begin{aligned} S(t) &= S(t_{(0)}) \prod_{j:t_{(j)} \leq t} [1 - P(t_{(j-1)} < T < t_{(j)} | T > t_{(j)})] \\ &= \prod_{j:t_{(j)} \leq t} [1 - q_j] \end{aligned}$$

Seja o estimador de q_j dado por:

$$\hat{q}_j = \frac{n^\circ \text{ de falhas em } t_{j-1}}{n^\circ \text{ de observações em risco em } t_{j-1}} = \frac{d_j}{n_j}$$

Teremos que o estimador de Kaplan-Meier será:

$$\hat{S}_{KM}(t) = \prod_{j:t_{(j)} \leq t} [1 - \hat{q}_j] = \prod_{j:t_{(j)} \leq t} [1 - \frac{d_j}{n_j}], \quad j = 1, 2, \dots$$

Não é possível obter a função de risco a partir dessa estimativa, mas sim a função de risco acumulada. A partir da relação entre $S(t)$ e $H(t)$, dada por:

$$S(t) = \exp[-H(t)]$$

Podemos usar a estimação de Kaplan-Meier nessa função e obter:

$$\hat{H}_{KM}(t) = -\log[\hat{S}_{KM}(t)]$$

Existem outras técnicas de estimação, mas esta permanece sendo a mais utilizada na literatura pois, no mesmo trabalho em que apresentam essa estimação, Kaplan e Meier (1958) mostram que $\hat{S}_{KM}(t)$ é um estimador de máxima verossimilhança de $S(t)$. A vantagem em usar estimadores dessa classe consiste nas boas propriedades assintóticas que possuem, consistentes e assintoticamente eficientes (PORTUGAL, 1995), tal como, por exemplo, convergência para uma distribuição Normal à medida que o tamanho da amostra aumenta. Com isso, assim como na estimação paramétrica, é possível obter intervalos de confiança e testes de hipóteses para estas estimativas.

3.2.4 Modelo de regressão de Cox

Afim de entender a relação entre o tempo até o evento de interesse e as covariáveis que têm influência na variação desse tempo, podemos utilizar de modelos de regressão com base na distribuição dos dados. No caso de modelos paramétricos, é necessário que se encontre a distribuição de probabilidade a qual a curva de sobrevivência melhor se ajusta, antes de propor um modelo de regressão. Já num modelo semi-paramétrico, esse ajuste não é necessário, o que facilita sua utilização desde que seus pressupostos sejam atendidos.

Como apresenta Colosimo e Giolo (2006), o modelo de Cox é um modelo semi-paramétrico que tem como pressuposto a proporcionalidade das taxas de falha entre os níveis de uma variável. Se considerarmos um modelo com apenas uma covariável dicotômica, onde a taxa de falha dos grupos 0 e 1 é, respectivamente, $\lambda_0(t)$ e $\lambda_1(t)$, e que existe proporcionalidade entre estas, temos que:

$$\frac{\lambda_1(t)}{\lambda_0(t)} = K$$

Onde K é a razão das taxas de falha, constante para qualquer tempo (t).

A partir disso, podemos compor o modelo de Cox. Se considerarmos x uma variável indicadora de grupo, onde:

$$x = \begin{cases} 0 & \text{se grupo 0} \\ 1 & \text{se grupo 1} \end{cases}$$

E $K = \exp(\beta x)$, podemos escrever:

$$\lambda(t|x) = \lambda_0(t) \exp(\beta x)$$

Essa função é o modelo de Cox com uma covariável. Se assumirmos p covariáveis e $\mathbf{x} = (x_1, \dots, x_p)'$ vetor de componentes, podemos encontrar a expressão geral desse modelo:

$$\lambda(t|\mathbf{x}) = \lambda_0(t)g(\mathbf{x}'\beta)$$

em que $\lambda_0(t)$ é o componente não-paramétrico, comumente chamado de taxa de falha de base, e $g(\mathbf{x}'\beta)$ é o componente paramétrico. Trata-se de uma função não negativa a ser especificada, frequentemente utilizada como:

$$g(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta) = \exp(\beta_1 x_1 + \dots + \beta_p x_p)$$

Esses betas são os valores dos parâmetros associados às covariáveis, garantindo nesse formato que $\lambda(t|\mathbf{x})$ não seja negativa. Nesse modelo vemos também que não há a presença de β_0 , uma vez que $\lambda_0(t)$ absorve esse termo constante.

Um dos principais pressupostos desse modelo é a proporcionalidade das taxas de falha. Ou seja, devemos ter que a razão da taxa de falhas K entre diferentes indivíduos será constante ao longo do tempo. Essa proporcionalidade pode ser avaliada pelo método gráfico, através do gráfico da função de sobrevivência estimada com o ajuste do modelo segundo uma covariável. Analisamos se as curvas para cada grupo são paralelas e, caso não sejam, podemos assumir que a variável não tem riscos proporcionais.

Os gráficos descritivos da estimativa de sobrevivência podem nos dar indícios desse comportamento, que serão confirmados com os gráficos ajustados pelo modelo e/ou com testes de hipóteses que testem a diferença das curvas dos grupos. Esses testes têm como hipóteses:

$$\left\{ \begin{array}{l} H_0 : S_a(t) = S_b(t) \text{ (Não existe diferença entre curvas de sobrevivência)} \\ H_1 : S_a(t) \neq S_b(t) \text{ (Existe diferença entre curvas de sobrevivência)} \end{array} \right.$$

Alguns dos mais famosos testes são Mantel-Haenszel, LogRank, Wilcoxon e família de testes, que é uma generalização de vários testes. Devemos nos atentar que os dois primeiros pressupõe que os riscos relativos das tabelas de contingência são iguais, ou seja, admitem que há proporcionalidade dos riscos. O teste LogRank tem como pressuposto essa característica, enquanto o teste de Wilcoxon não assume esse fator e por isso pode ser utilizado quando não há proporcionalidade dos riscos para verificar diferenças entre curvas.

Ao ser observada a proporcionalidade da covariável, podemos considerá-la no modelo. Caso contrário, não devemos incluí-la para evitar problemas de vício na estimação dos coeficientes do modelo (COLOSIMO; GIOLO, 2006 apud STRUTHERS; KALBFLEISCH, 1986).

Para escolha do melhor modelo, foram utilizadas a Máxima Verossimilhança Parcial (MVP) e o Teste de Razão de Verossimilhanças (TRV). A Máxima Verossimilhança Parcial utiliza a mesma ideia da Máxima Verossimilhança, com a intenção de obter-se estimadores com boas propriedades assintóticas que facilitarão a construção de testes de hipóteses e intervalos de confiança, mas com a possibilidade de trabalhar com censuras. Já o TRV nos permite fazer uma comparação entre modelos com ou sem alguma variável e o modelo completo, testando se β_i tem inclusão pertinente ou não para o modelo.

Para avaliar a qualidade do ajuste do modelo, utilizamos a proposição de Cox e Snell (1968) para análise de resíduos afim de avaliar o ajuste do modelo de Cox. Nesse modelo, os chamados resíduos de Cox-Snell são definidos por:

$$\hat{e}_i = \hat{\Lambda}_0(t_i) \exp \left[\sum_{k=1}^p x_{ik} \hat{\beta}_k \right], \quad i = 1, \dots, n$$

onde \hat{e}_i são os resíduos do modelo e $\hat{\Lambda}_0(t_i)$ é o estimador de sobrevivência de Nelson-Aalen, semelhante à Kaplan-Meier, descrito por:

$$\hat{\Lambda}_0(t_i) = \sum_{j:t_j < t} \left(\frac{d_j}{n_j} \right)$$

Entendemos que um modelo com bom ajuste terá seus \hat{e}_i 's com comportamento próximo de uma distribuição exponencial padrão, cujo gráfico é aproximadamente uma reta. Caso o gráfico não apresente esse comportamento, assumimos que o modelo não está bem ajustado. No entanto, nesse caso não é possível determinar o tipo de problema que o ajuste possui, por isso utilizamos apenas para avaliar o ajuste e não a proporcionalidade das taxas de falha (COLOSIMO; GIOLO, 2006).

Também é possível fazer esta análise com os resíduos de Schoenfeld, obtidos a partir da diferença entre observações de covariáveis do tempo até a ocorrência do evento de interesse de um indivíduo e os valores esperados para o tempo, dado um determinado grupo de risco. No entanto, sua análise envolve tanto vetores de resíduos quanto covariáveis ajustadas pelo modelo, além de diferente definição de tempo para cada uma. Assim, demos preferência aos resíduos de Cox-Snell pela praticidade de sua análise.

4 Metodologia

4.1 Bancos de Dados

Para analisar a mortalidade infantil, serão utilizados os dados do SINASC e do SIM. Ambos são organizados e disponibilizados pelo DATASUS, unidade de informação de saúde do SUS, ligada ao Ministério da Saúde. Esses dados são coletados diuturnamente dentro das unidades de saúde pública pelos próprios funcionários do estabelecimento e, posteriormente, repassados ao Ministério. A partir disso, são organizadas em bancos de dados anuais e disponibilizados anualmente no portal do DATASUS.

4.2 Enfoque da pesquisa e preparação das bases

Serão estudados os indivíduos que nasceram entre 2013 e 2014, afim de identificar se vieram a óbito entre 2013 e 2015, antes de completarem 1 ano de idade. Para isso, foram selecionados nos bancos do SIM os óbitos cuja a idade fosse de 1 (para um minuto de vida) a 311 (11 meses), incluindo também 400 (indivíduo com menos de 1 ano, mas sem idade exata). Através da variável número da declaração de nascimento (NUMERODN), presente em ambos os bancos, foi feito o cruzamento entre as bases para unir informações sobre natalidade e mortalidade do indivíduo. Foi feito um *match* determinístico com os dados do banco SINASC para as observações do banco SIM que não continham esse registro.

Esse método consiste em comparar os valores de um grupo de variáveis para identificar se há correspondência entre as observações dos dois bancos, afim de definir se representam o mesmo indivíduo. As variáveis escolhidas para esse processo foram data de nascimento (DTNASC), sexo do indivíduo (SEXO), idade da mãe na data de ocorrência do parto (IDADEMAE), peso ao nascer (PESO), número de semanas de gestação (SEMAGESTAC), tipo de gravidez (GRAVIDEZ), tipo de parto (PARTO), código do estabelecimento de ocorrência (CODESTAB), bairro de residência (BAIRES), código do bairro de residência (CODBAIRES) e endereço de residência (ENDRES).

O batimento utilizando CODESTAB foi feito apenas nos casos em que o óbito ocorreu em menos de 30 dias. Consideramos igualdade suficiente nos casos em que DTNASC, SEXO, IDADEMAE, SEMAGESTAC, PESO e ENDRES eram iguais. Nos casos em que ENDRES não era correspondente, as outras variáveis foram analisadas e o *match* foi feito apenas no caso de total igualdade entre os valores das variáveis nos dois bancos de dados. Algumas observações apresentaram ausência de valores nessas variáveis que impossibilitou o *match*.

Após unir os bancos, realizamos uma seleção de variáveis para eliminar aquelas que estavam presentes em ambos SINASC e SIM, mantendo apenas uma no banco final. Foi dada preferência para as variáveis provenientes do SINASC, pois continham menos informações ausentes. Ainda assim, fizemos um segundo *match*, agora para os valores das variáveis escolhidas com valores ausentes. Aquelas cuja a variável correspondente no SIM continham informação assumiram esse valor. Aquelas cujo ambas as variáveis comparadas estavam como *missing* permaneceram assim.

Um caso especial desse *match* foi a variável hora do óbito (HORAOBITO). Foram analisadas as datas de nascimento e óbito (DTNASC e DTOBITO), hora do nascimento (HORANASC) e idade do indivíduo na ocasião do óbito (IDADE) para as observações de HORAOBITO com *missing*. Os indivíduos que faleceram antes de completarem 24h de vida tiveram o número de horas presente na IDADE acrescidas ao horário de nascimento, sem alterar os minutos. Para os indivíduos que faleceram após 1 dia de vida, assumimos o mesmo valor de HORANASC para HORAOBITO.

Aqueles que não faleceram até 1 ano de idade foram acrescidos posteriormente, apenas com as informações de natalidade, sendo tratados como censura na análise, formando assim o banco a ser utilizado para análise. Por fim, foram retiradas desse banco final as observações que estavam com *missing* para uma ou mais variáveis dentre peso ao nascer (PESONASC), APGAR de 1º minuto (APGAR1), APGAR de 5º minuto (APGAR5), presença de anomalia congênita ou má formação (IDANOMAL), nº de consultas pré natais (CONSULTAS), tempo (em semanas) de gestação (GESTACAO), tipo de gravidez (GRAVIDEZ), tipo de parto (PARTO), raça/cor da mãe (RACACORMAE), estado civil da mãe (ESTCIVMAE), escolaridade da mãe (ESCMAE), ocupação da mãe (OCUPMAE), nº de filhos tidos vivos (QTDFILVIVO), nº de filhos tidos mortos (QTDFILMORT), nº de partos normais tidos pela mãe (QTDPARTNOR) e nº de partos cesáreos tidos pela mãe (QTDPARTCES). A variável raça/cor da criança foi retirada nessa limpeza. Apesar de ser uma variável importante em uma análise, o alto número de *missings* torna a variável pouco informativa.

Para realizar a Análise de sobrevivência são necessárias as variáveis que apresentem o tempo decorrido até o indivíduo sofrer o evento de interesse (óbito antes de 1 ano de idade) e a variável indicadora de censura, ou seja, se o indivíduo foi exposto ou não ao evento. Como essas variáveis não estão presentes no banco, elas foram criadas a partir daquelas já existentes.

A variável tempo até o óbito (tempo) foi criada a partir da diferença entre a junção das variáveis data e hora de nascimento e a junção da data e hora de óbito, com valor em horas. Para aqueles que não vieram a óbito foi assumido o valor de 8760 horas, referente a 12 meses completos de idade, determinado como fim do tempo de observação. Como algumas observações apresentaram tempo igual a 0, foi acrescida uma fração de 0,01 hora

em todas as observações.

Já a variável indicadora de censura (censura) atribuiu valor 1 aos indivíduos que vieram a óbito antes de completarem 1 ano de idade, indicada pela data de óbito, e valor 0 àqueles que não sofreram o evento de interesse, cuja data de óbito apresentava *missing*.

4.3 Técnica

Aplicamos a Análise de Sobrevivência, técnica com base em análise de regressão em que a variável-resposta para o modelo é o tempo de falha, ou seja, o tempo transcorrido desde o início da observação do fenômeno até a ocorrência de falha ou censura. A censura é a observação parcial de uma resposta, ou seja, observações que por alguma razão tiveram seu acompanhamento interrompido. Essa característica configura-se a principal característica de dados de sobrevivência, pois essa técnica, diferente das outras, permite que sejam utilizadas as informações de uma observação censurada (COLOSIMO; GIOLO, 2006).

Foi trabalhado um modelo onde a variável de interesse para a Análise de Sobrevivência será o tempo até o óbito do indivíduo antes do primeiro ano de vida. Indivíduos que não foram identificados nos registro de mortalidade ou que seu falecimento tenha ocorrido após completar 1 ano de idade foram considerados censuras. Para as funções de sobrevivência e taxa de falha, utilizamos a estimação de Kaplan Meier, conhecida e muito utilizada pela literatura por ser um estimador de máxima verossimilhança da função de sobrevivência.

Foi utilizado o modelo de regressão de Cox para avaliar a influência das covariáveis na função taxa de falha, ou seja, encontrar um modelo que explique o tempo até o óbito. Para escolha do modelo mais adequado, foi feita uma seleção de variáveis manual onde foram gerados todos os modelos possíveis com as variáveis pré-selecionadas de acordo com a análise da proporcionalidade dos riscos. A qualidade do ajuste será avaliada através do gráficos dos resíduos de Cox-Snell.

Para a análise de dados, a aplicação da técnica e a modelagem foram utilizados os softwares R e Excel.

5 Resultados

Nos anos de 2013 a 2015, nasceram 89209 crianças residentes no Distrito Federal, das quais 882 vieram a óbito antes de completarem 1 ano de idade (evento de interesse), de acordo com os registros do SINASC e SIM. Para esta análise foram consideradas apenas as observações cuja as variáveis não apresentaram *missings*. Assim, compõem esta análise as observações de 47321 nascimentos, dentre os quais 419 sofreram o evento de interesse.

5.1 Análise descritiva dos dados

5.1.1 Análise descritiva clássica

As variáveis escolhidas para compor a análise foram divididas em três categorias: características da criança, características gestacionais e do parto e características maternas. Analisando o primeiro grupo segundo a variável censura, podemos ver que não há muita diferença na distribuição de sexo, tanto entre indivíduos censurados quanto entre os não censurados.

A variável APGAR traz o sistema de *scores* proposto por Apgar (1952). Ela propõe uma escala para determinar rápida e eficientemente as condições de um recém-nascido, numa tentativa de prever a chance de sobrevivência da criança. Avalia-se 5 características: aparência, pulso, gesticulação, atividade e respiração, cuja nota é dada no 1º e no 5º minutos de vida da criança. A escala varia de 0 a 10, em que se classifica asfixia grave (com apgar entre 0 e 2), asfixia moderada (entre 3 e 4), asfixia leve (entre 5 e 7) e sem asfixia (entre 8 e 10). Entre os indivíduos censurados, podemos notar que a grande maioria não apresenta asfixia tanto no 1º (88,8%) quanto no 5º minuto (98,4%). Já os indivíduos que vieram a óbito tiveram classificação mais distribuída. Ainda que a maioria não tenha apresentado asfixia nas duas análises (35,1% e 57,7%, respectivamente), há uma porcentagem razoável de indivíduos com asfixia grave no 1º (19,3%) e no 5º minutos (12%).

Com relação ao peso ao nascer, 65,9% dos indivíduos censurados tinham 3kg ou mais, enquanto apenas 15% daqueles não censurados apresentaram a mesma classificação. A maioria destes nasceu com peso entre 500g e 1,9kg. Por fim, com relação à presença de anomalia ou má formação, apenas 0,5% dos indivíduos censurados apresentaram essa condição, contra 15,8% daqueles que sofreram o evento.

Tabela 1 – Características da criança segundo presença de censura

Variável	Indivíduos censurados	Indivíduos não censurados
Sexo		
Feminino	49,2%	48,7%
Masculino	50,8%	51,3%
APGAR 1º minuto		
Sem asfixia	88,8%	35,1%
Asfixia leve	9,8%	30,4%
Asfixia moderada	1,0%	15,3%
Asfixia grave	0,4%	19,3%
APGAR 5º minuto		
Sem asfixia	98,4%	57,7%
Asfixia leve	1,4%	23,7%
Asfixia moderada	0,1%	6,6%
Asfixia grave	0,1%	12,0%
Peso ao nascer		
Menos de 500g	0,05%	5,0%
Entre 500g e 999g	0,2%	32,0%
Entre 1kg e 1,9kg	2,4%	23,6%
Entre 2kg e 2,9kg	31,5%	24,3%
3kg ou mais	65,9%	15,0%
Presença de anomalia		
Apresenta	0,5%	15,8%
Não apresenta	99,5%	84,2%

Tabela 2 – Características gestacionais e de parto segundo presença de censura

Variável	Indivíduos censurados	Indivíduos não censurados
Tipo de gravidez		
Única	97,7%	91,6%
Dupla	2,3%	7,6%
Tripla e mais	0,04%	0,7%
Tempo da gestação		
Menos de 22 semanas	0,1%	1,2%
22 a 27 semanas	1,1%	29,8%
28 a 31 semanas	3,8%	16,5%
32 a 36 semanas	45,0%	20,0%
37 a 41 semanas	41,2%	31,5%
41 semanas ou mais	8,8%	1,0%
Nº de consultas pré-natal		
Nenhuma	2,6%	9,8%
De 1 a 3	5,8%	21,7%
De 4 a 6	20,5%	31,3%
7 ou mais	71,1%	37,2%
Tipo de parto		
Vaginal	41,3%	41,8%
Cesáreo	58,7%	58,2%

Tabela 3 – Características maternas segundo presença de censura

Variável	Indivíduos censurados	Indivíduos não censurados
Idade		
Menos de 14 anos	0,6%	1,2%
Entre 15 e 19 anos	11,9%	13,4%
Entre 20 e 29 anos	43,9%	45,1%
Entre 30 e 39 anos	40,0%	33,7%
40 anos ou mais	3,5%	6,7%
Raça/cor		
Branca	31,9%	30,1%
Preta	3,9%	4,8%
Amarela	0,6%	0,5%
Parda	63,5%	63,7%
Indígena	0,2%	1,0%
Escolaridade (em anos)		
Nenhuma	0,1%	1,4%
1 a 3 anos	1,5%	1,7%
4 a 7 anos	13,4%	18,6%
8 a 11 anos	49,1%	47,5%
12 anos ou mais	35,9%	30,8%
Estado Civil		
Solteira	31,2%	4,7%
Casada	40,6%	54,8%
Viúva	0,1%	0,4%
Separada judicialmente	1,2%	2,5%
Divorciada	26,9%	37,6%
Nº de filhos vivos		
Nenhum	46,1%	47,7%
1 filho	32,5%	27,9%
2 filhos	13,3%	12,2%
3 a 5 filhos	7,5%	11,9%
6 filhos ou mais	0,7%	0,2%
Nº de filhos mortos		
Nenhum	81,8%	78,0%
1 filho	14,6%	14,6%
2 filhos	2,7%	5,0%
3 a 5 filhos	0,8%	2,4%
6 filhos ou mais	0,04%	0,0%
Nº de partos normais anteriores		
Nenhum	68,6%	67,3%
De 1 a 3	28,8%	28,6%
De 4 a 6	2,3%	3,8%
7 ou mais	0,3%	0,2%
Nº de partos cesáreos anteriores		
Nenhum	72,8%	74,9%
De 1 a 3	27,0%	24,6%
De 4 a 6	0,2%	0,5%
7 ou mais	0,03%	0,0%

Dentre as características da gestação e do parto, foi possível observar que mais de 90% dos indivíduos entre ambos são provenientes de uma gravidez única. Já o tempo da gestação tem maioria concentrada entre 32 e 41 semanas para indivíduos que sobreviveram.

Mas para aqueles que vieram a óbito não há uma concentração, com 31,5% das gestações durando entre 37 e 41 semanas, seguido por 29,8% com duração entre 22 e 27 semanas.

A maioria dos indivíduos censurados teve um maior acompanhamento pré-natal, denotando 7 ou mais consultas pré-natais, enquanto indivíduos não censurados ficaram mais distribuídos entre as categorias, ainda que a maioria (37,2%) tenha tido 7 ou mais consultas. Com relação ao parto, a distribuição foi parecida, com maioria de partos cesáreos.

O grupo de variáveis que informam as características da mãe apresentou os resultados mais parecidos para indivíduos censurados e não censurados, com porcentagens muito próximas entre as categorias de cada uma das variáveis.

5.1.2 Análise descritiva da sobrevivência

Com base nas variáveis tempo e censura, obtemos a curva de sobrevivência estimada por Kaplan-Meier para óbitos de indivíduos com menos de 1 ano de idade, em horas. Pode-se observar na Figura 1 que a probabilidade de sobrevivência à esse evento é muito elevada, mantendo-se acima de 99% até os 12 meses incompletos de observação. A taxa de risco (Figura 2) apresenta um crescimento muito rápido antes de 625 horas de vida (em torno de 26 dias) e vai se estabilizando até o fim do período observado.

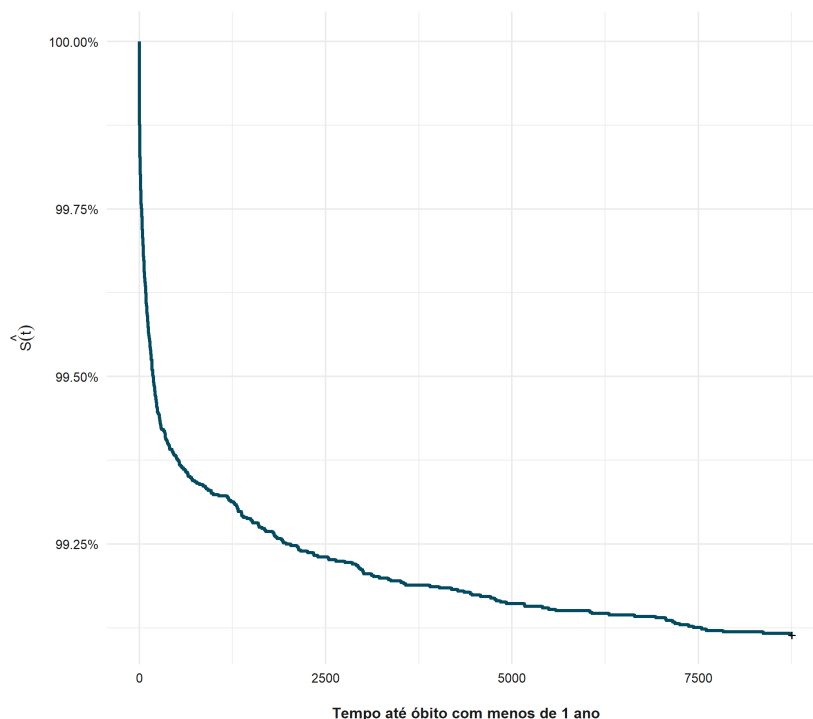


Figura 1 – Tempo de sobrevivência do indivíduo até o óbito (em horas)

Investigamos em seguida o comportamento da sobrevivência segundo as covariáveis presentes no banco de dados para saber se essas influenciam nesse resultado. Seguimos a mesma abordagem anterior, analisando os grupos de características da criança, da

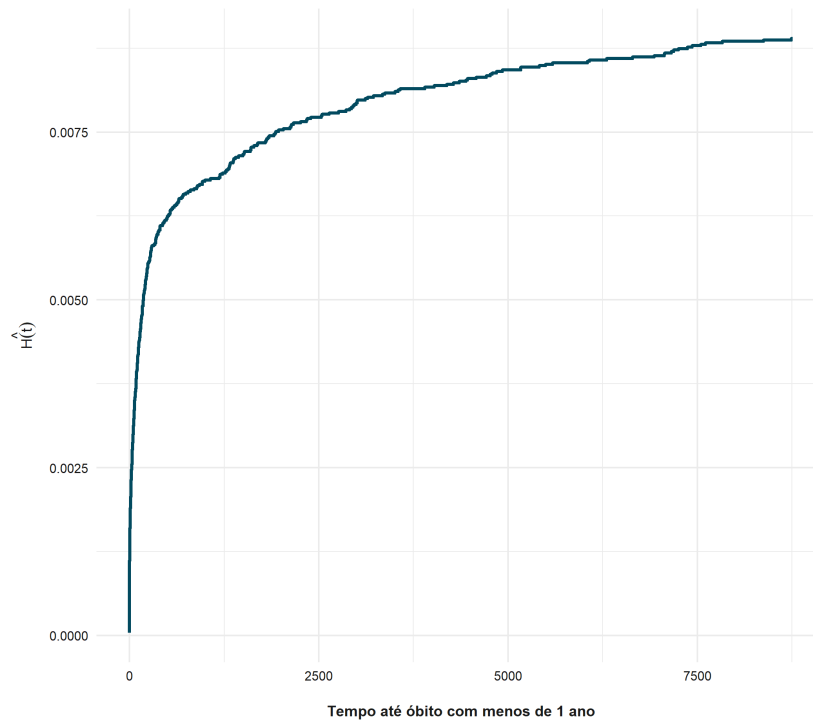


Figura 2 – Taxa de risco acumulado para ocorrência do óbito

gestação e da mãe. Curvas com comportamentos diferentes para níveis diferentes são o comportamento de interesse a ser observado nesses gráficos.

Além disso, para aplicar a modelagem de regressão de Cox, devemos verificar o pressuposto de proporcionalidade dos riscos segundo as covariáveis em análise. Para isso, avaliamos se as curvas são paralelas entre os níveis ao longo do tempo em conjunto os resultados com o teste de Wilcoxon para ter indícios de atendimento ao pressuposto.¹

Desta forma, denotamos que entre as características infantis, as variáveis APGAR de 1º e 5º minutos, peso ao nascer e presença de anomalia apresentaram diferença clara entre os níveis das respectivas variáveis. No caso da presença de anomalia, por exemplo, a probabilidade de sobrevivência de um indivíduo que apresente algum tipo de anomalia ao nascer decai ao longo do primeiro ano de vida para menos de 75%. A diferença entre as curvas dos níveis para as três variáveis é confirmada pelo teste de Wilcoxon, a nível de significância de 1%. No entanto, apenas APGAR de 1º minuto (Figura 3) e presença de anomalia (Figura 4) apresentaram proporcionalidade de acordo com suas curvas para poderem atender ao modelo.

Quanto às características gestacionais e de parto, foi possível notar que, com exceção do tipo de parto, as variáveis apresentam razoável distinção na probabilidade de sobrevivência entre seus níveis. Por Wilcoxon, confirmamos que existem evidências para rejeitar a hipótese de não haver diferença nas curvas, com 1% de significância. A gravidez

¹ São apresentados ao longo do texto os gráficos que atendem à diferença de curvas e ao pressuposto de Cox. Os gráficos restantes podem ser encontrados no Apêndice

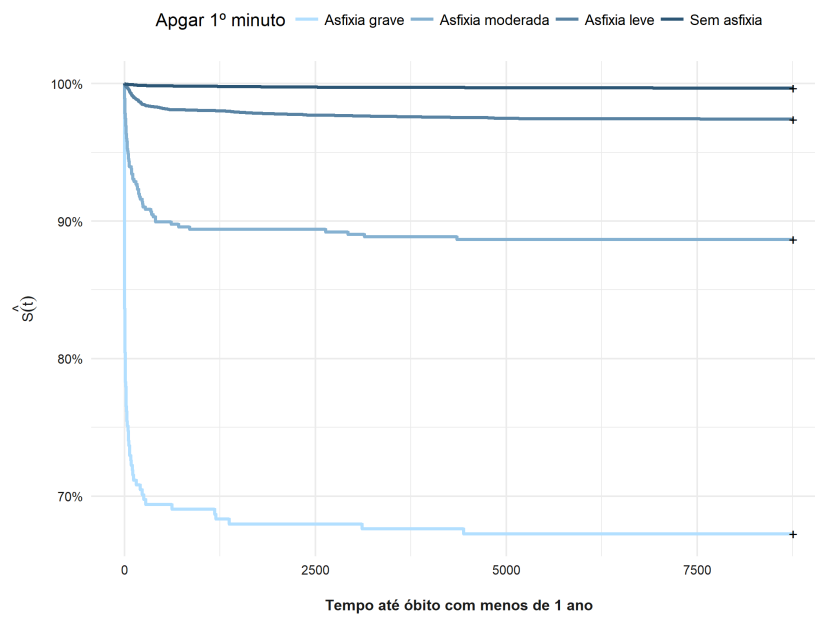


Figura 3 – Tempo de sobrevivência segundo Apgar de 1º minuto.

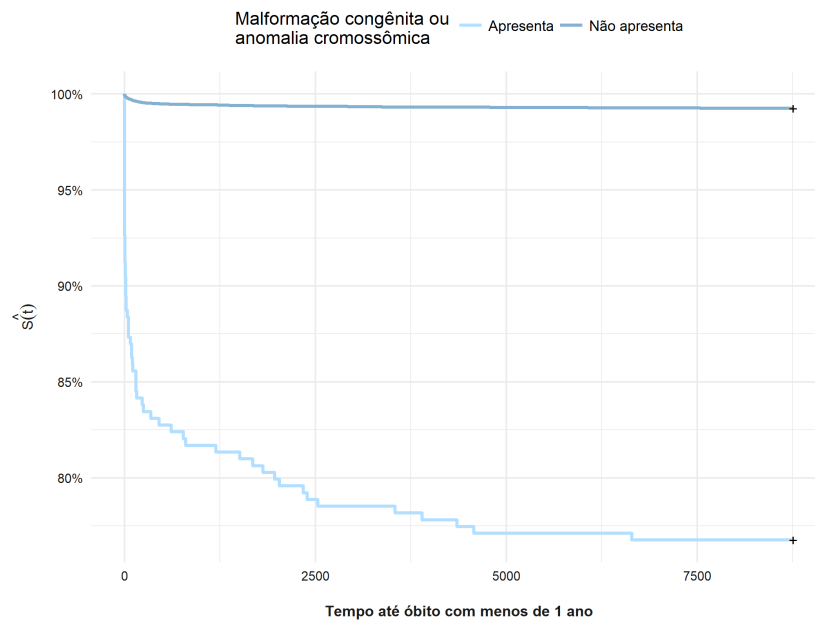


Figura 4 – Tempo de sobrevivência segundo presença de anomalia ou malformação congênita.

de trigêmeos ou mais, por exemplo, apresenta maior risco de óbito do que a de um filho único ou de gêmeos, uma vez que a probabilidade de sobrevivência é de pouco mais de 85% (Figura 5). Esta foi a única variável que apresentou curvas paralelas entre seus níveis, atendendo ao pressuposto da modelagem proposta.

Com relação às características da mãe, notou-se diferenças nos padrões das curvas para cada nível. No entanto, as probabilidades estão muito próximas entre si. Para confirmar

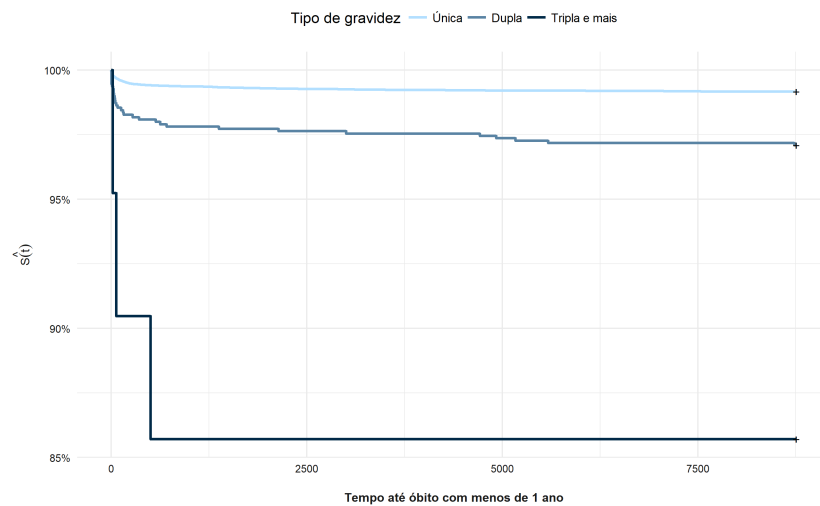


Figura 5 – Tempo de sobrevivência segundo tipo de gravidez.

se existia ou não diferença entre as curvas, realizamos o teste de Wilcoxon. Com p-valor menor que 0,01, podemos dizer termos evidências de que as variáveis de idade, raça/cor, escolaridade, número de filhos vivos e número de filhos mortos apresentam diferença nas curvas de sobrevivência de seus respectivos níveis. Já as variáveis estado civil, número de partos normais e número de partos cesáreos não apresentam evidências de diferença entre suas curvas, a um nível de 5% de significância. Por fim, apenas a variável número de filhos mortos apresentou indícios de proporcionalidade das curvas.

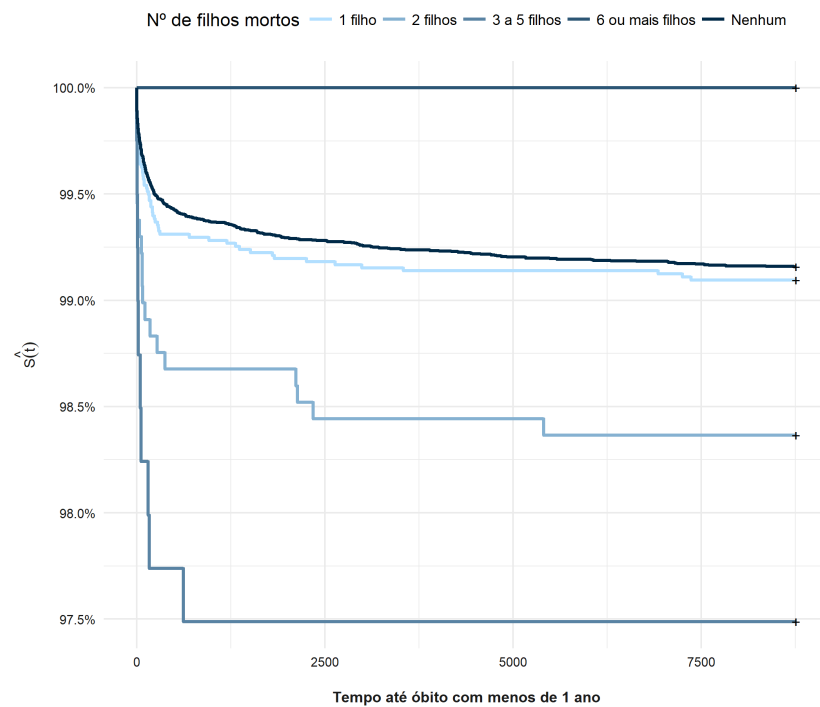


Figura 6 – Tempo de sobrevivência segundo número de filhos tidos mortos.

5.2 Aplicação do modelo de regressão de Cox

Com os resultados da seção anterior, tem-se que apenas as variáveis APGAR de 1º minuto, presença de anomalia, tipo de gravidez, número de filhos tidos mortos e número de partos cesáreos anteriores apresentaram indícios de proporcionalidade dos riscos e puderam ser consideradas para inclusão no modelo.

Uma vez que estas variáveis são categorizadas, será necessário criar variáveis *dummy* para que essas se adequem corretamente ao modelo. Para uma variável com k níveis, devemos criar $k - 1$ *dummys*. Como exemplo, temos as categorias sem asfixia, asfixia leve, asfixia moderada e asfixia grave. Portanto, sendo a asfixia grave a categoria de referência, as *dummys* para APGAR1 serão:

$$D_{i,1} = \begin{cases} 1, & \text{se não apresenta asfixia} \\ 0, & \text{caso contrário} \end{cases}$$

$$D_{i,2} = \begin{cases} 1, & \text{se apresenta asfixia leve} \\ 0, & \text{caso contrário} \end{cases}$$

$$D_{i,3} = \begin{cases} 1, & \text{se apresenta asfixia moderada} \\ 0, & \text{caso contrário} \end{cases}$$

Para encontrar o melhor modelo, foram construídos todos os modelos possíveis com as variáveis pré-selecionadas. Considere V1: APGAR 1º minuto, V2: presença de anomalia, V3: número de filhos tidos mortos, V4: tipo de gravidez e V5: quantidade de partos cesáreos anteriores. Esses modelos foram avaliados a partir do valor da máxima verossimilhança parcial e do Teste de Razão de Verossimilhanças (TRV) (Tabela 4). Para MVP, o modelo de referência para o cálculo é o modelo sem covariáveis (nulo).

Com base apenas nessas medidas, não foi possível definir por um modelo apenas, uma vez que vários apresentaram convergência no TRV e valores de verossimilhança parcial bastante próximos. Então, reavaliamos os modelos, agora levando em consideração o p-valor das variáveis *dummy* e o R^2 calculado para cada modelo ².

Dois modelos resultaram boa resposta: o modelo com APGAR de 1º minuto (APGAR1) e presença de anomalia (IDANOMAL) (V1+V2) e o modelo com APGAR de 1º minuto, presença de anomalia e tipo de gravidez (GRAVIDEZ) (V1+V2+V4). Dentre todos, esses foram os únicos que tiveram p-valor significativo para todas as variáveis

² Ver Apêndice B

Tabela 4 – Seleção de Covariáveis através da modelagem de Cox

Passos	Modelo	$-2 \log L(\theta)$	TRV	p-valor
Passo 1	Nulo (Referência)	-5184	-	-
	V1	-4435	1173	<2e-16
	V2	-4660	365	<2e-16
	V3	-5177	14	0,006
	V4	-5159	50	1,00E-11
	V5	-5183	2	0,5
Passo 2	V1+V2	-4039	1282	<2e-16
	V1+V3	-4431	1181	<2e-16
	V1+V4	-4421	1201	<2e-16
	V1+V5	-4434	1174	<2e-16
	V2+V3	-4652	380	<2e-16
	V2+V4	-4637	410	<2e-16
	V2+V5	-4658	368	<2e-16
	V3+V4	-5152	63	9,00E-12
	V3+V5	-5176	17	0,02
	V4+V5	-5158	53	4,00E-10
Passo 3	V1+V2+V3	-4036	1288	<2e-16
	V1+V2+V4	-4025	1311	<2e-16
	V1+V2+V5	-4037	1287	<2e-16
	V1+V3+V4	-4417	1209	<2e-16
	V1+V3+V5	-4430	1183	<2e-16
	V1+V4+V5	-4420	1203	<2e-16
	V2+V3+V4	-4631	422	<2e-16
	V2+V3+V5	-4651	383	<2e-16
	V2+V4+V5	-4636	413	<2e-16
	V3+V4+V5	-5151	66	<2e-16
Passo 4	V1+V2+V3+V4	-4022	1316	<2e-16
	V1+V2+V3+V5	-4034	1294	<2e-16
	V1+V2+V4+V5	-4023	1315	<2e-16
	V1+V3+V4+V5	-4416	1210	<2e-16
	V2+V3+V4+V5	-4630	425	<2e-16
Passo 5	V1+V2+V3+V4+V5	-4020	1321	<2e-16

dummy, ao nível de 99% de confiança. Há ainda uma diferença mínima entre os valores de R^2 . Para o primeiro, o R^2 resultou em 0,03. No segundo tivemos $R^2 = 0,027$.

Desta forma, investigamos primeiro as curvas de sobrevivência estimadas a partir dos parâmetros β de Cox. Escolhemos a análise gráfica devido ao tamanho da amostra ($n = 47571$), o que leva a p-valores muito pequenos em testes de hipóteses. Comparamos as curvas dos níveis das variáveis com o *baseline*, ou linha de base, que serve como referência. Para o modelo com APGAR (*baseline* = Asfixia grave) e IDANOMAL (*baseline* = apresenta anomalia), vemos nas Figuras 7 e 8 que as curvas são paralelas com relação à linha de base e o tempo de sobrevivência é diferente para os níveis ao longo do período observado. Isso confirma a proporcionalidade do risco e a influência das variáveis na sobrevivência, justificando sua presença no modelo

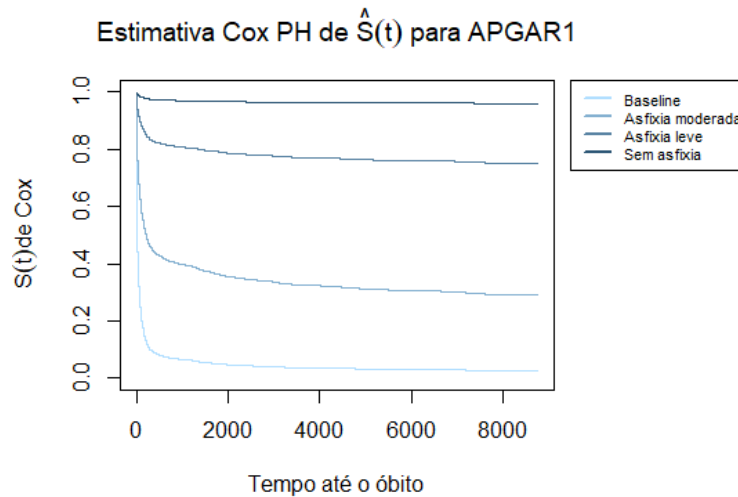


Figura 7 – Estimativa Cox-PH do modelo V1+V2 para $\hat{S}(t)$ segundo APGAR 1

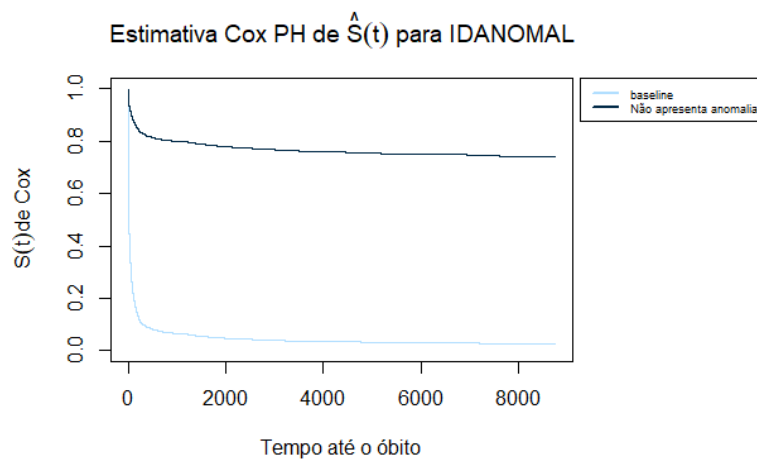


Figura 8 – Estimativa Cox-PH do modelo V1+V2 para $\hat{S}(t)$ segundo APGAR 1

Já o modelo com as variáveis APGAR, IDANOMAL e GRAVIDEZ apresenta um comportamento de $\hat{S}(t)$ um pouco diferente do que vimos na pré-análise. Em especial, a variável GRAVIDEZ (*baseline* = Tripla ou mais) apresenta encontro entre o *baseline* e a variável *dummy* que representa a gravidez do tipo dupla. Com isso, não podemos garantir o pressuposto de proporcionalidade de Cox para essa variável e, ainda que seja influente, não podemos incluí-la no modelo final.

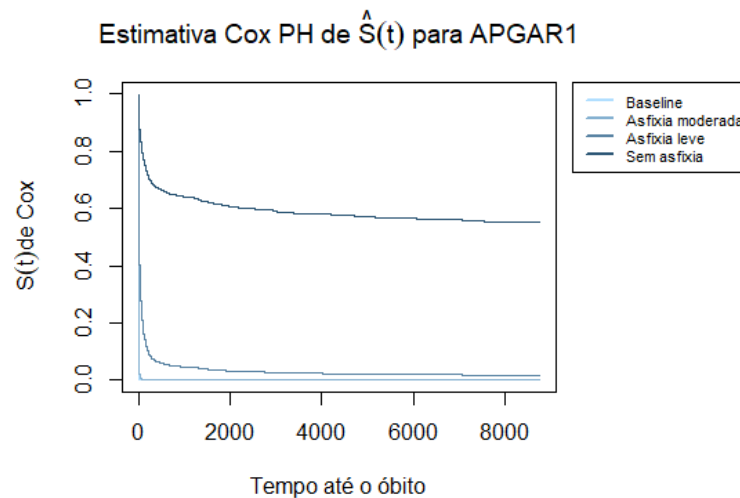


Figura 9 – Estimativa Cox-PH do modelo V1+V2+V4 para $\hat{S}(t)$ segundo APGAR 1

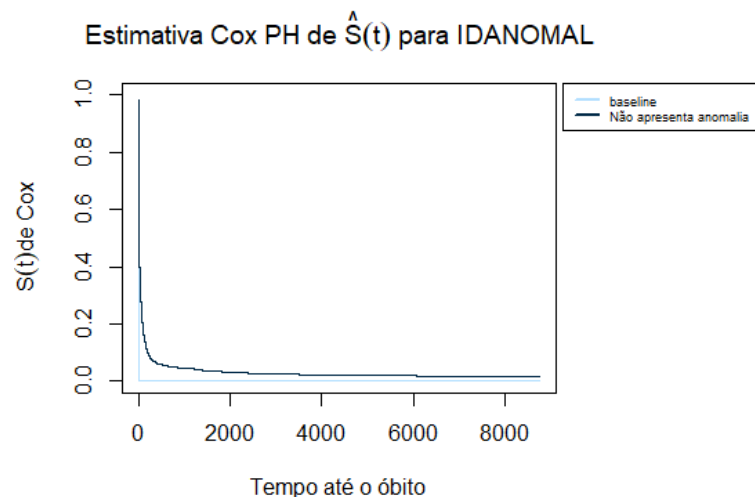


Figura 10 – Estimativa Cox-PH do modelo V1+V2+V4 para $\hat{S}(t)$ segundo IDANOMAL

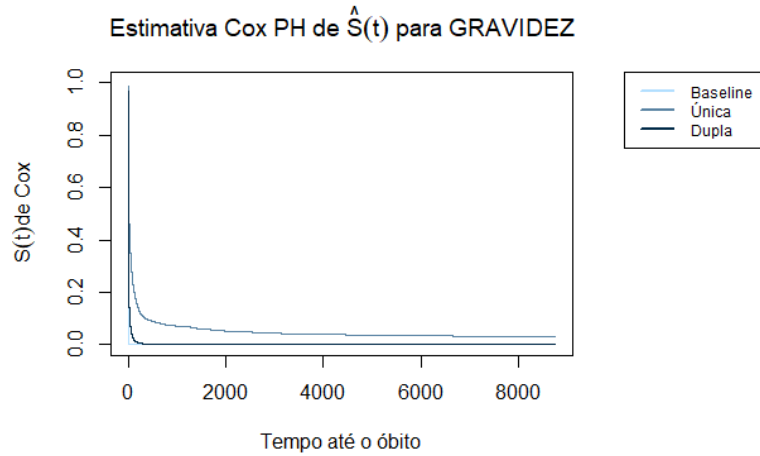


Figura 11 – Estimativa Cox-PH do modelo V1+V2+V4 para $\hat{S}(t)$ segundo GRAVIDEZ

Desta forma, opta-se por utilizar o modelo que contem APGAR e IDANOMAL, dado que se adequa à modelagem de Cox. A expressão do modelo é dada por:

$$\lambda(t|\underline{x}) = \lambda_0(t) \exp(\hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4)$$

E suas estimativas estão na Tabela 5. Para interpretar os coeficientes, devemos considerar a propriedade de taxas de falha proporcionais do modelo. Calculando a razão das taxas de falha de dois indivíduos, i e j , com mesmos valores para as covariáveis exceto da l -ésima, temos que:

$$\frac{\lambda(t|x_i)}{\lambda(t|x_j)} = \exp[\beta(x_{il} - x_{jl})]$$

Mas, como todas as variáveis deste modelo são dicotômicas, a taxa de óbito entre um nível será $\exp(\beta)$ vezes o nível de base. No caso da variável APGAR, o nível de base é a asfixia grave. Portanto, temos que a taxa de morte de crianças com asfixia moderada é 67,6% menor que a de crianças com asfixia grave. Os outros níveis apresentam resultado semelhante, com a taxa de morte por asfixia leve 92% menor que por asfixia grave e sem asfixia 99% menor.

A variável IDANOMAL tem como nível de base a presença de anomalia. Portanto, podemos dizer que a taxa de morte entre indivíduos não apresentam anomalia é 92% menor do que entre os que apresentam. Esses resultados de ambas variáveis corroboram com o que podemos ver nos gráficos de tempo de sobrevivência estimada.

Tabela 5 – Estimativas do modelo de Cox

APGAR1+IDANOMAL	$\hat{\beta}$	$\exp(\hat{\beta})$	$EP(\hat{\beta})$	z	$Pr(> z)$	$IC_{95\%} \exp(\hat{\beta})$
APGAR1: Asfixia moderada	-1.08	0.34	0.17	-6.52	6.92e-11	(0,024;0,468)
APGAR1: Asfixia leve	-2.55	0.08	0.14	-17.99	< 2e-16	(0,059;0,103)
APGAR1: sem asfixia	-4.51	0.01	0.14	-32.13	< 2e-16	(0,008;0,014)
IDANOMAL: Não apresenta	-2.50	0.08	0.14	-17.66	< 2e-16	(0,062;0,108)

Comparamos o modelo obtido com os modelos mais simples, apenas com APGAR ou com IDANOMAL, através da Análise de Variância. O teste que é realizado se baseia na medida qui-quadrado (χ^2) e testa se a redução na soma de quadrados dos resíduos do modelo é estatisticamente significativa entre um modelo e outro. Seja o resíduo de um modelo denotado por ϵ , as hipóteses do teste são:

$$\begin{cases} \sum_i \epsilon_{i,A}^2 = \sum_i \epsilon_{i,B}^2 \\ \sum_i \epsilon_{i,A}^2 > \sum_i \epsilon_{i,B}^2 \end{cases}$$

A partir disso, podemos definir se a inclusão de uma segunda variável é boa para o modelo. Vemos nas tabelas 6 e 7 que o p-valor é muito baixo, temos evidências para rejeitar a hipótese nula do teste. Assim, podemos afirmar que temos indícios que a inclusão de IDANOMAL no modelo com apenas APGAR ou de APGAR no modelo com apenas IDANOMAL contribui para o modelo. Vemos que é uma adição positiva pois a Log-Verossimilhança parcial aumenta, um indício de que o segundo modelo é melhor que o primeiro, em ambos os casos.

Tabela 6 – Análise de Variância comparando os modelos APGAR e APGAR+IDANOMAL

	loglik	χ^2	Df	$P(> \chi)$
1	-4025.38			
2	-3929.38	192.00	1	0.0000

Tabela 7 – Análise de Variância comparando os modelos IDANOMAL e APGAR+IDANOMAL

	loglik	χ^2	Df	$P(> \chi)$
1	-4376.72			
2	-3929.38	894.68	3	0.0000

Analizamos também os resíduos de Cox-Snell para a sobrevivência ajustada no modelo de Cox, através da Figura 12. Podemos perceber que os resíduos estimados se distanciam bastante da curva exponencial padrão ajustada pelos dados, formato que deveriam ter caso o modelo estivesse bem ajustado. Entendemos então que o modelo de Cox proposto não se ajusta bem aos dados, apesar de ter tido o melhor resultado entre os outros propostos.

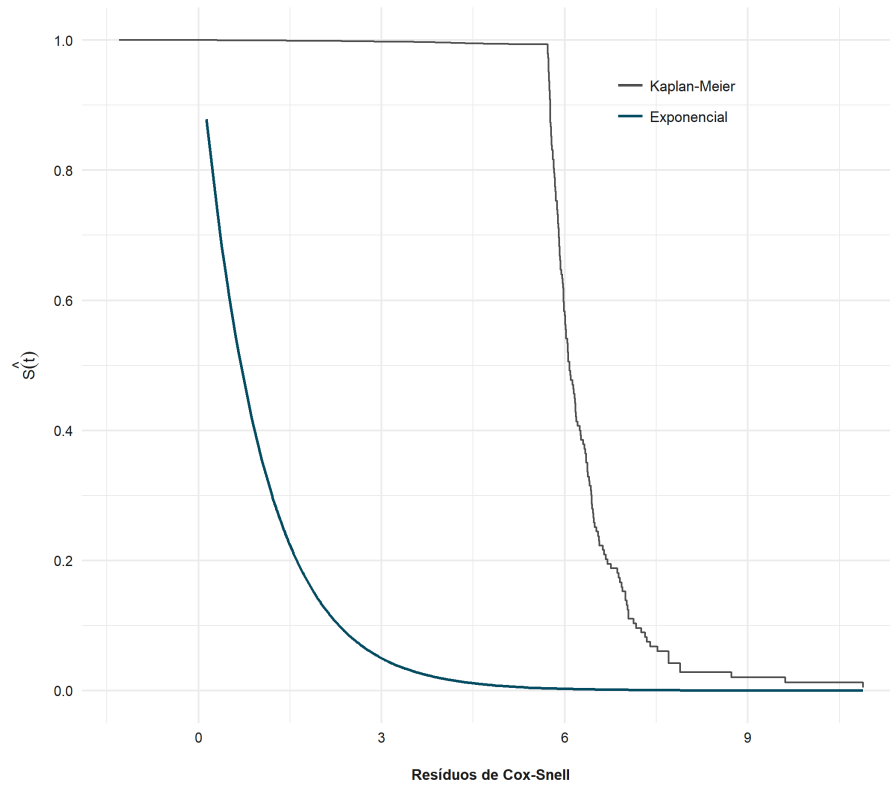


Figura 12 – Resíduos de Cox-Snell para $\hat{S}(t)$

6 Conclusões

Ainda que suas taxas tenham caído drasticamente nas últimas décadas, a mortalidade infantil no Distrito Federal ainda deve receber atenção. Mesmo com baixos índices, devemos entender porque elas ainda ocorrem e quais circunstâncias levam esses indivíduos ao óbito. Neste trabalho, discutimos que o padrão de mortalidade neonatal e pós neonatal no DF seguiu o padrão nacional, com alta queda da TMI a partir dos anos 80.

Procuramos também explicar a técnica de Análise de Sobrevivência e sua aplicabilidade. Vimos que este é um tipo de estudo bastante relevante quando a variável de interesse do estudo é o tempo transcorrido até um determinado evento. É possível analisar também a relação desta com as covariáveis em um conjunto de dados, o que nos ajuda a entender e explicar o tempo estudado. A partir disso, definimos como evento de interesse óbito com menos de 1 ano de idade de residentes no Distrito Federal, para os nascidos entre 2013 e 2014. Mostramos como se comportam os tempos de sobrevivência e a taxa de risco acumulado do com relação ao tempo até a ocorrência do evento.

Aqui tivemos uma resposta positiva: mais de 99% dos indivíduos sobrevivem ao primeiro ano de vida. No entanto, notamos também que são nos primeiros 26 dias de vida que a taxa de risco acumulada tem seu crescimento intensificado e a probabilidade de sobrevivência é reduzida de maneira mais evidente. Isso pode nos indicar que a atenção neonatal precisa ser intensificada afim de diminuir mais ainda os casos de óbitos infantis. Analisamos também quais características do bebê, de sua gestação e de sua mãe poderiam estar relacionadas com a ocorrência de seu óbito em menos de 1 ano completo, utilizando o modelo de regressão de Cox.

Primeiramente, analisamos descritivamente as variáveis relacionadas ao nascimento de indivíduos que sofreram ou não o evento de interesse. Estudamos o comportamento da sobrevivência com relação a cada uma das covariáveis, afim de fazer uma pré-avaliação de quais destas entrariam ou não no modelo. Essa análise foi feita para garantir a proporcionalidade dos riscos das variáveis, pressuposto que deve ser garantido para aplicação desta modelagem. Os indícios da análise gráfica foram confirmados utilizando o teste de Wilcoxon para verificação de igualdade das curvas de sobrevivência.

Com isso, 5 variáveis mostraram atender esse pressuposto: apgar de 1º minuto, presença de anomalia, número de filhos mortos tidos, tipo de gravidez e número de partos cesáreos anteriores. Foram construídos todos modelos possíveis para esse conjunto de variáveis e avaliou-se a partir de alguns índices qual era o modelo mais explicativo para os dados estudados.

O modelo final é composto por duas variáveis, APGAR de 1º minuto (APGAR1) e

presença de anomalia ou malformação congênita (IDANOMAL). Comparamos esse modelo com o modelo com apenas uma variável e notou-se que há ganhos em utilizar o modelo escolhido, uma vez que este tem a maior capacidade de descrição dos dados.

O modelo de Cox foi escolhido para esse trabalho por ser um modelo bastante aplicado em estudos na área da saúde, especialmente em estudos clínicos (COLOSIMO; GIOLO, 2006). No entanto, observou-se que muitas variáveis não atendiam seu pressuposto básico e por isso não puderam entrar no modelo, ainda que graficamente apresentassem influência nos tempos de sobrevivência dos indivíduos. Embora o modelo tenha tido ajuste razoável, é pouco explicativo (com $R^2 = 0,03$), o que pode ser resultado da pouca informação que apenas duas variáveis fornecem ao modelo estimado. Além disso, vimos na análise de resíduos que o modelo escolhido pelo melhor resultado quanto à Máxima Verossimilhança Parcial, ao TRV e ao p-valor das variáveis não têm uma boa qualidade de ajuste.

Desta forma, concluímos que o modelo de Cox não se mostrou ideal esse conjunto de dados e, por isso, seria interessante explorar outras modelagens em estudos futuros. Uma direção seriam os modelos paramétricos, cujo ajuste não necessita do pressuposto de proporcionalidade dos riscos. Assim, variáveis como peso ao nascer, tempo de gestação, número de consultas pré-natais, entre outras, que sugerem por meio das curvas de sobrevivência terem influência no tempo até o evento de interesse mas não atendem ao pressuposto do modelo de Cox poderiam ser incluídas, afim de encontrar um modelo mais explicativo e melhor ajustado ao conjunto de dados.

Outra abordagem seria a utilização de modelos de fração de cura. Vimos que mais de 99% dos indivíduos não vieram a óbito, ou seja, não sofreram o evento de interesse. Em Análise de Sobrevivência, esses são ditos indivíduos "curados- nomeclatura proveniente da aplicação em casos clínicos. Com o alto índice de "cura"desse fenômeno, seria interessante utilizar modelos desse tipo, visando melhorar ainda mais o ajuste e a explicabilidade do fenômeno em estudo.

Referências

- APGAR, V. A proposal for a new method of evaluation of the newborn. *Classic Papers in Critical Care*, Springer, v. 32, n. 449, p. 97, 1952. Citado na página 39.
- BARROS, F. C. et al. Mortalidade perinatal e infantil em pelotas, rio grande do sul: nossas estatísticas são confiáveis? *Cadernos de Saúde Pública*, SciELO Public Health, v. 1, n. 3, p. 348–358, 1985. Citado 2 vezes nas páginas 17 e 21.
- BASTOS, J.; ROCHA, C. Análise de sobrevivência métodos não paramétricos. *Arquivos de Medicina*, Arquimed-Departamento de Edições Científicas da AEFMUP, v. 21, n. 3-4, p. 111–114, 2007. Citado na página 29.
- BELIZZI, A. Características da mortalidade infantil na coorte de nascimentos de 1992 no município de niterói (rio de janeiro)[dissertação de mestrado]. *Rio de Janeiro: Instituto de Medicina Social da Universidade Estadual do Rio de Janeiro*, 1996. Citado na página 21.
- CAMPOS, T. P.; CARVALHO, M. S.; BARCELLOS, C. C. Mortalidade infantil no rio de janeiro, brasil: áreas de risco e trajetória dos pacientes até os serviços de saúde. *Revista Panamericana de Salud Publica*, SciELO Public Health, v. 8, p. 164–171, 2000. Citado na página 21.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Edgar Blücher, 2006. Citado 8 vezes nas páginas 23, 27, 28, 31, 32, 33, 37 e 54.
- COSTA, M. d. C. N. et al. Mortalidade infantil no brasil em períodos recentes de crise econômica. *Revista de Saúde Pública*, SciELO Public Health, v. 37, p. 699–706, 2003. Citado 2 vezes nas páginas 17 e 21.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 248–275, 1968. Citado na página 32.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Citado na página 30.
- MAGALHÃES, M. N. *Probabilidade e variáveis aleatórias*. [S.l.]: Edusp, 2006. Citado na página 25.
- MONTEIRO, R. A.; SCHMITZ, B. d. A. S. Principais causas básicas da mortalidade infantil no distrito federal, brasil: 1990 a 2000. Instituto de Medicina Integral Prof. Fernando Figueira, 2004. Citado 2 vezes nas páginas 17 e 22.
- MOSLEY, W. H.; CHEN, L. C. An analytical framework for the study of child survival in developing countries. *Population and development review*, JSTOR, v. 10, p. 25–45, 1984. Citado 2 vezes nas páginas 21 e 22.
- NAKANO, E. Y. Um curso de análise de sobrevivência. Universidade de Brasília. 2017. Citado 2 vezes nas páginas 27 e 30.

- NAKANO, E. Y.; CARRASCO, C. G. Uma avaliação do uso de um modelo contínuo na análise de dados discretos de sobrevivência. *Trends in Applied and Computational Mathematics*, v. 7, n. 1, p. 91–100, 2006. Citado na página 25.
- PAIM, J. S.; DIAS, C. N.; ARAÚJO, J. D. d. Influência de fatores sociais e ambientais na mortalidade infantil. 1980. Citado na página 21.
- PORTUGAL, M. S. Notas introdutórias sobre o princípio de máxima verossimilhança: Estimação e teste de hipóteses. *UFRGS, RS*, n. 1986, p. 1–24, 1995. Citado na página 30.
- SECRETARIA DE SAÚDE DO DISTRITO FEDERAL. *Boletim Epidemiológico - Mortalidade Infantil no Distrito Federal*. [S.l.], 2015. Citado 2 vezes nas páginas 17 e 22.
- STRUTHERS, C. A.; KALBFLEISCH, J. D. Misspecified proportional hazard models. *Biometrika*, Oxford University Press, v. 73, n. 2, p. 363–369, 1986. Citado na página 32.
- SZWARCWALD, C. L. et al. Mortalidade infantil no brasil: Belíndia ou bulgária? *Cadernos de Saúde Pública*, SciELO Public Health, v. 13, n. 3, p. 503–516, 1997. Citado 2 vezes nas páginas 17 e 21.
- THE WORLD BANK. *Social Indicators of Development 1995*. [S.l.], 1995. Citado na página 21.
- VASCONCELOS, A. M. N.; GOMES, M. M. F. Transição demográfica: a experiência brasileira. *Epidemiologia e Serviços de Saúde*, Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços/Secretaria de Vigilância em Saúde/Ministério da Saúde, v. 21, n. 4, p. 539–548, 2012. Citado na página 21.

APÊNDICE A – Gráficos de Sobrevivência segundo covariáveis

A.1 Tempos de sobrevivência segundo características da criança

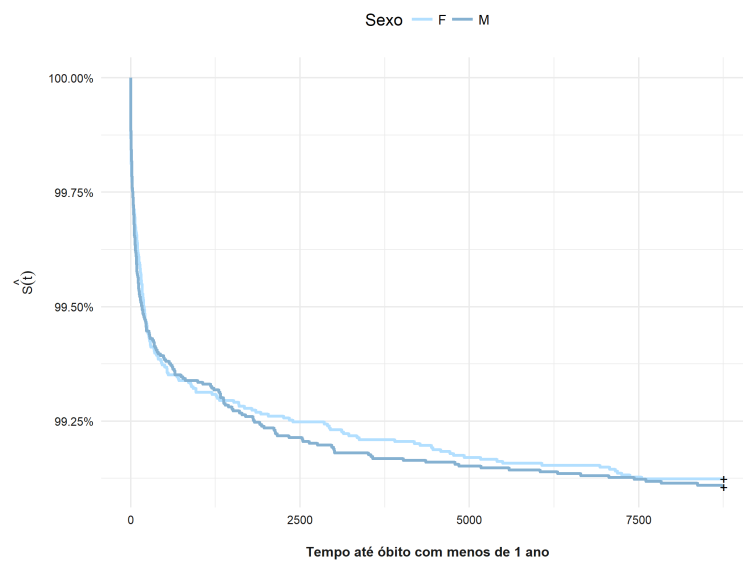


Figura 13 – Tempo de sobrevivência do indivíduo segundo sexo.

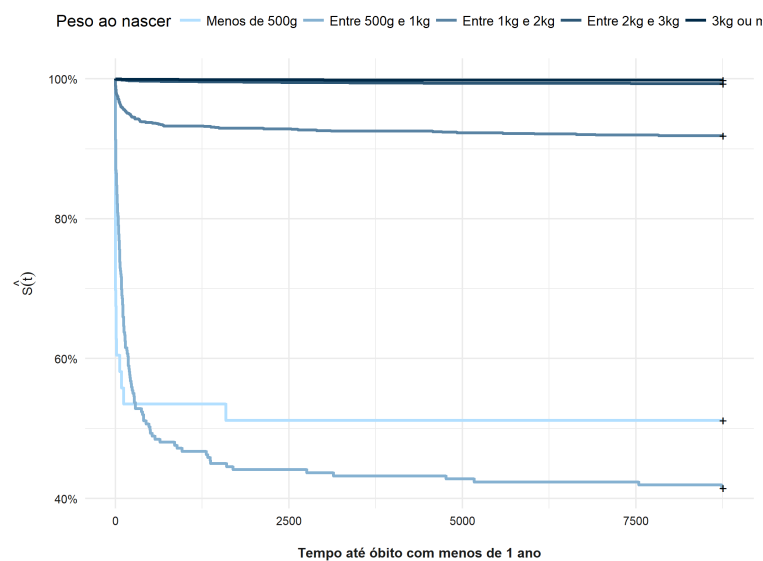


Figura 14 – Tempo de sobrevivência segundo peso ao nascer.

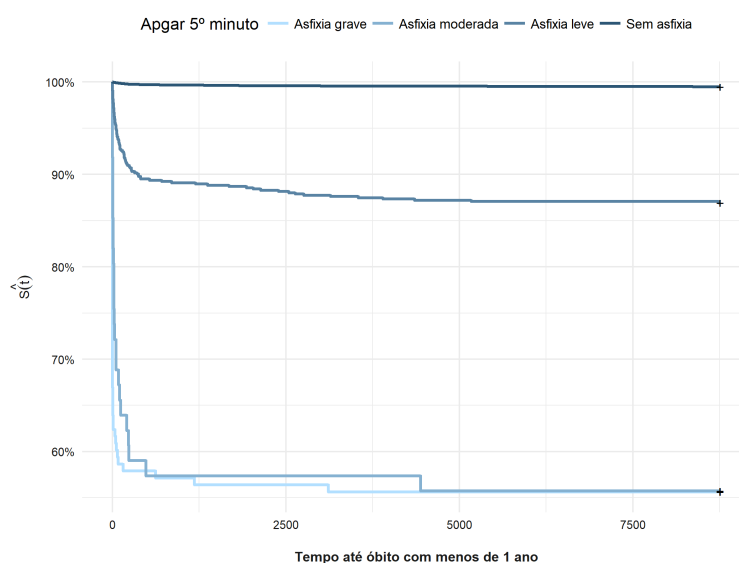


Figura 15 – Tempo de sobrevivência segundo APGAR de 5º minuto.

A.2 Tempos de sobrevivência segundo características gestacionais e de parto

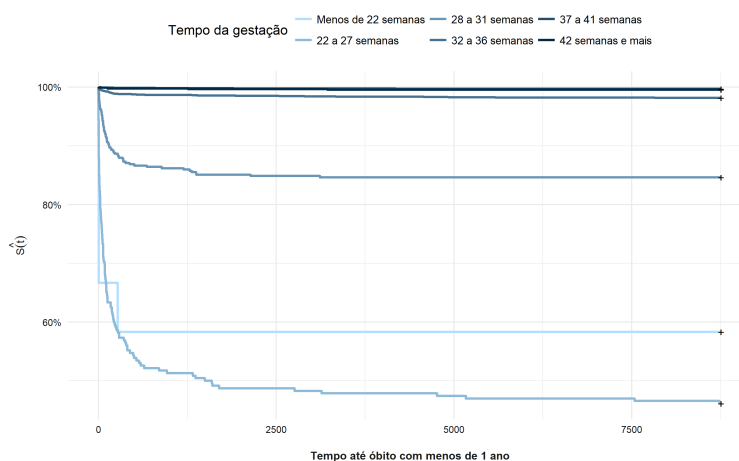


Figura 16 – Tempo de sobrevivência segundo tempo de gestação.

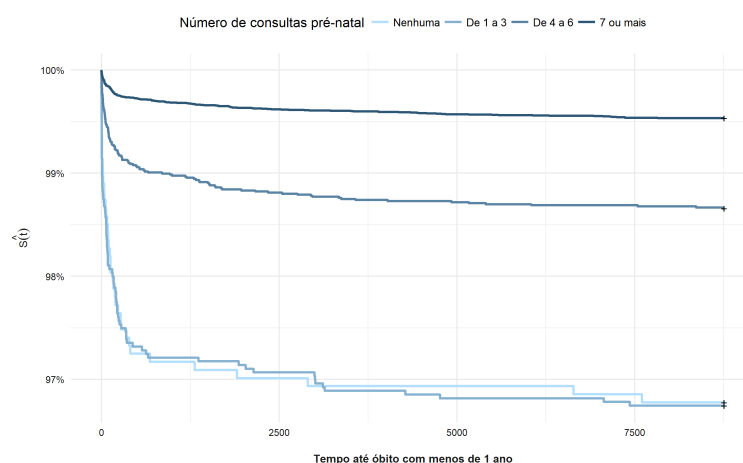


Figura 17 – Tempo de sobrevivência segundo nº de consultas pré-natais.

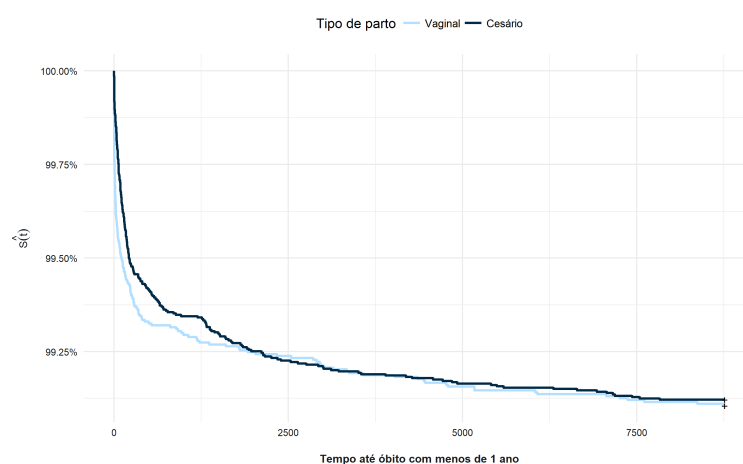


Figura 18 – Tempo de sobrevivência segundo tipo de parto.

A.3 Tempos de sobrevivência segundo características maternas

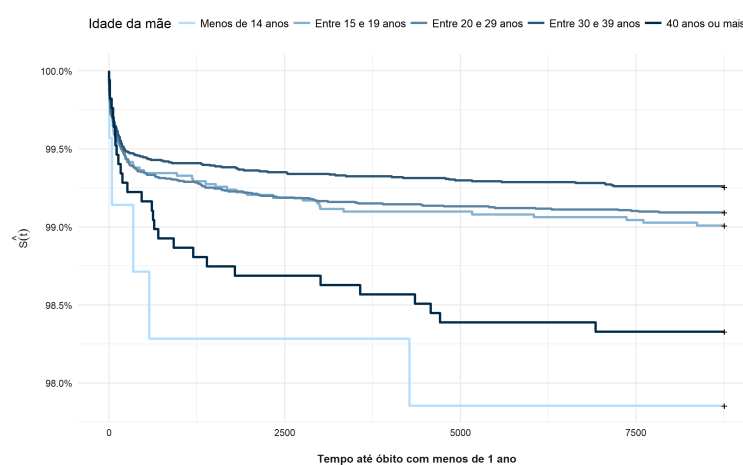


Figura 19 – Tempo de sobrevivência segundo idade da mãe.

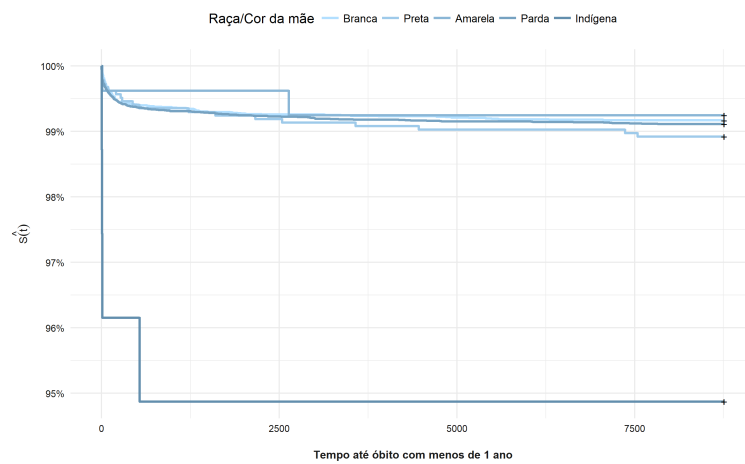


Figura 20 – Tempo de sobrevivência segundo raça/cor da mãe.

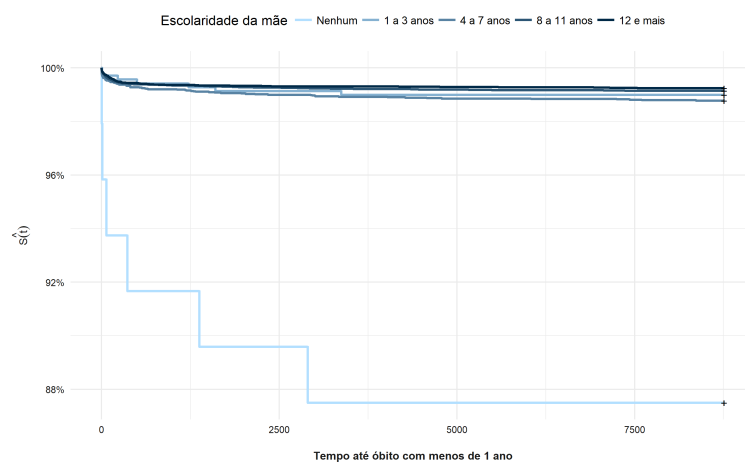


Figura 21 – Tempo de sobrevivência segundo escolaridade da mãe.

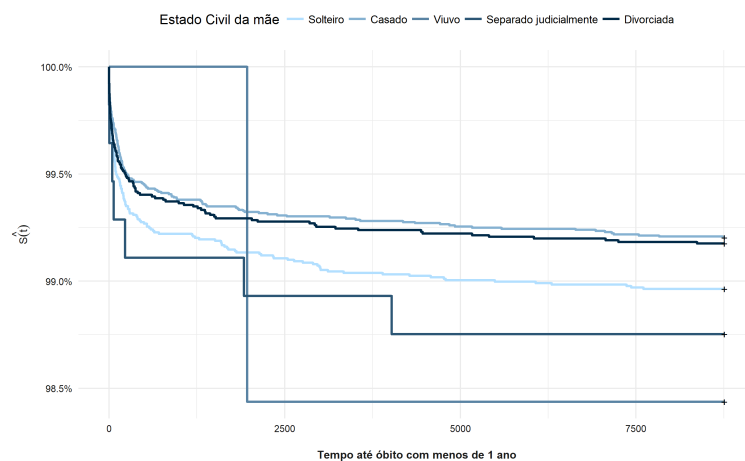


Figura 22 – Tempo de sobrevivência segundo estado civil da mãe.

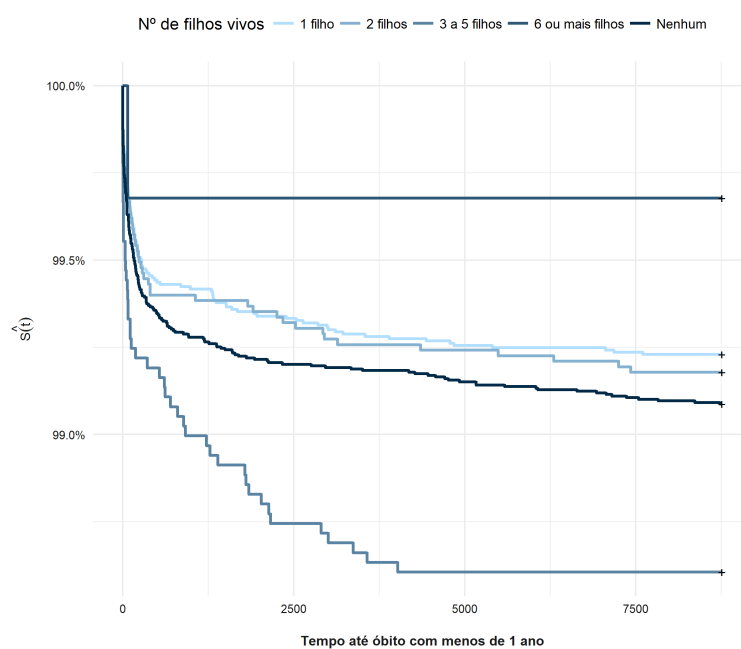


Figura 23 – Tempo de sobrevivência segundo número de filhos tidos vivos.

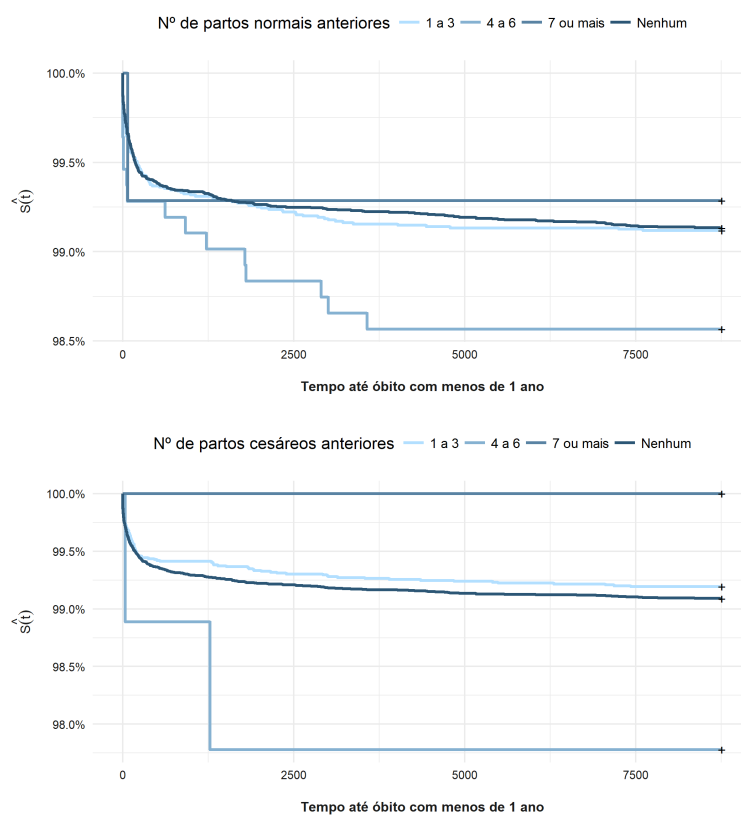


Figura 24 – Tempo de sobrevivência segundo nº de partos tidos anteriormente (normais e/ou cesáreos).

APÊNDICE B – Modelos de Cox

Tabela 8 – Coeficientes do modelo estimado completo

Modelo completo	β	$EP(\beta)$	$\exp(\beta)$	\mathbf{z}	$Pr(> z)$	R^2
APGAR1: Asfixia moderada	-1,13	0,32	0,16	-6,87	6,36E-12	0,027
APGAR1: Asfixia leve	-2,59	0,07	0,14	-18,26	<2e-16	
APGAR1: Sem asfixia	-4,53	0,01	0,14	-32,07	<2e-16	
IDANOMAL: Não apresenta	-2,6	0,07	0,14	-18,7	<2e-16	
QTDFILMORT: 2 filhos	0,14	1,15	0,25	0,58	0,56	
QTDFILMORT: 3 a 5 filhos	0,94	2,55	0,38	2,48	0,01	
QTDFILMORT: 6 ou mais filhos	-12,03	5,97E-06	1154	-0,01	0,99	
QTDFILMORT: Nenhum	0,01	1,02	0,14	0,11	0,92	
GRAVIDEZ: Única	-2,59	0,08	0,58	-4,45	8,55E-06	
GRAVIDEZ: Dupla	-1,72	0,18	0,61	-2,84	0,004	
QTDPARTCES: 4 a 6	1,29	3,61	0,72	1,79	0,07	
QTDPARTCES: 7 ou mais	-13,26	1,75E-06	1722	-0,01	0,99	
QTDPARTCES: Nenhum	0,17	1,19	0,12	1,48	0,14	

Tabela 9 – Coeficientes dos modelos estimados com 2 variáveis

Modelo	β	$EP(\beta)$	$\exp(\beta)$	\mathbf{z}	$Pr(> z)$	R^2
APGAR1+QTDFILMORT						
APGAR1: Asfixia moderada	-1,29	0,28	0,15	-8,4	<2e-16	0,02
APGAR1: Asfixia leve	-2,92	0,05	0,13	-22	<2e-16	
APGAR1: Sem asfixia	-4,93	0,01	0,13	-38,28	<2e-16	
QTDFILMORT: 2 filhos	0,33	1,39	0,24	1,38	0,17	
QTDFILMORT: 3 a 5 filhos	0,98	2,67	0,37	2,62	0,01	
QTDFILMORT: 6 ou mais filhos	-11,77	0	905,6	-0,01	0,99	
QTDFILMORT: Nenhum	-0,03	0,97	0,13	-0,2	0,84	
APGAR1+GRAVIDEZ						
APGAR1: Asfixia moderada	-1,31	0,27	0,15	-8,52	<2e-16	0,02
APGAR1: Asfixia leve	-2,93	0,05	0,13	-22,17	<2e-16	
APGAR1: Sem asfixia	-4,91	0,01	0,13	-38,31	<2e-16	
GRAVIDEZ: Única	-2,4	0,09	0,58	-4,13	3,62E-05	
GRAVIDEZ: Dupla	-1,52	0,22	0,6	-2,52	0,01	
APGAR1+QTDPARTCES						
APGAR1: Asfixia moderada	-1,29	0,28	0,15	-8,43	<2e-16	0,02
APGAR1: Asfixia leve	-2,92	0,05	0,13	-22,1	<2e-16	
APGAR1: Sem asfixia	-4,94	0,01	0,13	-38,45	<2e-16	
QTDPARTCES: 4 a 6	0,86	2,37	0,71	1,21	0,23	
QTDPARTCES: 7 ou mais	-11,56	0	642,9	-0,02	0,99	
QTDPARTCES: Nenhum	-0,01	0,99	0,11	-0,13	0,9	
IDANOMAL+QTDFILMORT						
IDANOMAL: Não apresenta	-3,58	0,03	0,13	-27,83	<2e-16	0,01
QTDFILMORT: 2 filhos	0,54	1,72	0,24	2,24	0,03	
QTDFILMORT: 3 a 5 filhos	0,97	2,63	0,34	2,86	0,004	
QTDFILMORT: 6 ou mais filhos	-11,18	0	683,1	-0,02	0,99	
QTDFILMORT: Nenhum	-0,08	0,93	0,13	-0,57	0,57	
IDANOMAL+GRAVIDEZ						
IDANOMAL: Não apresenta	-3,59	0,03	0,13	-27,85	<2e-16	0,01
GRAVIDEZ: Única	-2,95	0,05	0,58	-5,08	3,78E-07	

Tabela 9 continuação da página anterior

GRAVIDEZ: Dupla	-1,74	0,18	0,6	-2,88	0,004	
IDANOMAL+QTDPARTCES						
IDANOMAL: Não apresenta	-3,6	0,03	0,13	-27,94	<2e-16	0,01
QTDPARTCES: 4 a 6	1,05	2,85	0,71	1,47	0,14	
QTDPARTCES: 7 ou mais	-11,05	0	788,9	-0,01	0,99	
QTDPARTCES: Nenhum	0,11	1,11	0,11	0,96	0,34	
QTDFILMORT+GRAVIDEZ						
QTDFILMORT: 2 filhos	0,51	1,67	0,24	2,14	0,03	0
QTDFILMORT: 3 a 5 filhos	0,84	2,32	0,34	2,49	0,01	
QTDFILMORT: 6 ou mais filhos	-11,28	0	668,3	-0,02	0,99	
QTDFILMORT: Nenhum	-0,08	0,92	0,13	-0,66	0,51	
GRAVIDEZ: Única	-2,78	0,06	0,58	-4,81	1,55E-06	
GRAVIDEZ: Dupla	-1,53	0,22	0,6	-2,55	0,01	
QTDFILMORT+QTDPARTCES						
QTDFILMORT: 2 filhos	0,54	1,72	0,24	2,26	0,02	0
QTDFILMORT: 3 a 5 filhos	0,89	2,42	0,34	2,62	0,01	
QTDFILMORT: 6 ou mais filhos	-12,05	0	963,3	-0,01	0,99	
QTDFILMORT: Nenhum	-0,11	0,9	0,13	-0,86	0,39	
QTDPARTCES: 4 a 6	0,65	1,92	0,71	0,91	0,36	
QTDPARTCES: 7 ou mais	-11,95	0	1057	-0,01	0,99	
QTDPARTCES: Nenhum	0,15	1,16	0,11	1,4	0,16	
GRAVIDEZ+QTDPARTCES						
GRAVIDEZ: Única	-2,76	0,06	0,58	-4,76	1,93E-06	0
GRAVIDEZ: Dupla	-1,48	0,23	0,6	-2,46	0,01	
QTDPARTCES: 4 a 6	0,7	2,01	0,71	0,98	0,33	
QTDPARTCES: 7 ou mais	-11,23	0	721,6	-0,02	0,99	
QTDPARTCES: Nenhum	0,12	1,12	0,11	1,1	0,27	

Tabela 10 – Coeficientes dos modelos estimados com 3 variáveis

Modelo	β	$EP(\beta)$	$\exp(\beta)$	\mathbf{z}	$Pr(> z)$	R^2	
APGAR1+IDANOMAL+QTDFILMORT							
APGAR1: Asfixia moderada	-1,11	0,33	0,16	-6,77	1,29E-11	0,026	
APGAR1: Asfixia leve	-2,57	0,08	0,14	-18,07	<2e-16		
APGAR1: Sem asfixia	-4,55	0,01	0,14	-32,14	<2e-16		
IDANOMAL: Não apresenta	-2,55	0,08	0,14	-18,44	<2e-16		
QTDFILMORT: 2 filhos	0,1	1,1	0,25	0,4	0,69		
QTDFILMORT: 3 a 5 filhos	0,97	2,63	0,38	2,57	0,01		
QTDFILMORT: 6 ou mais filhos	-11,98	1,00E-05	1060	-0,01	0,99		
QTDFILMORT: Nenhum	0	1	0,14	-0,01	1		
APGAR1+IDANOMAL+GRAVIDEZ							
APGAR1: Asfixia moderada	-1,11	0,33	0,16	-6,85	7,24E-12	0,027	
APGAR1: Asfixia leve	-2,58	0,08	0,14	-18,35	<2e-16		
APGAR1: Sem asfixia	-4,52	0,01	0,14	-32,32	<2e-16		
IDANOMAL: Não apresenta	-2,58	0,08	0,14	-18,87	<2e-16		
GRAVIDEZ: Única	-2,57	0,08	0,58	-4,43	9,66E-06		
GRAVIDEZ: Dupla	-1,67	0,19	0,61	-2,76	0,0058		
APGAR1: Asfixia moderada	-1,11	0,33	0,16	-6,81	9,49E-12		0,026
APGAR1: Asfixia leve	-2,57	0,08	0,14	-18,33	<2e-16		
APGAR1: Sem asfixia	-4,54	0,01	0,14	-32,57	<2e-16		
IDANOMAL: Não apresenta	-2,6	0,07	0,14	-18,77	<2e-16		
QTDPARTCES: 4 a 6	1,4	4,05	0,72	1,95	0,05		
QTDPARTCES: 7 ou mais	-11,91	1,00E-05	908,2	-0,01	0,99		
QTDPARTCES: Nenhum	0,16	1,18	0,11	1,43	0,15		
APGAR1+QTDFILMORT+GRAVIDEZ							
APGAR1: Asfixia moderada	-1,31	0,27	0,15	-8,49	<2e-16	0,024	
APGAR1: Asfixia leve	-2,93	0,05	0,13	-22,07	<2e-16		
APGAR1: Sem asfixia	-4,91	0,01	0,13	-38,07	<2e-16		
QTDFILMORT: 2 filhos	0,31	1,36	0,24	1,27	0,2		
QTDFILMORT: 3 a 5 filhos	0,97	2,64	0,37	2,6	0,01		
QTDFILMORT: 6 ou mais filhos	-11,73	1,00E-05	916,1	-0,01	0,99		
QTDFILMORT: Nenhum	-0,01	0,99	0,13	-0,08	0,94		
GRAVIDEZ: Única	-2,42	0,09	0,58	-4,17	3,05E-05		
GRAVIDEZ: Dupla	-1,56	0,21	0,6	-2,59	0,01		
APGAR1+QTDFILMORT+QTDPARTCES							
APGAR1: Asfixia moderada	-1,29	0,28	0,15	-8,4	<2e-16	0,023	
APGAR1: Asfixia leve	-2,92	0,05	0,13	-21,98	<2e-16		
APGAR1: Sem asfixia	-4,93	0,01	0,13	-38,19	<2e-16		
QTDFILMORT: 2 filhos	0,33	1,39	0,24	1,38	0,17		
QTDFILMORT: 3 a 5 filhos	0,98	2,67	0,37	2,63	0,01		
QTDFILMORT: 6 ou mais filhos	-11,77	1,00E-05	908,7	-0,01	0,99		
QTDFILMORT: Nenhum	-0,03	0,97	0,13	-0,2	0,84		
QTDPARTCES: 4 a 6	0,87	2,38	0,71	1,22	0,22		
QTDPARTCES: 7 ou mais	-12,53	4,00E-06	1063	-0,01	0,99		
QTDPARTCES: Nenhum	0,01	1,01	0,11	0,1	0,92		
APGAR1+GRAVIDEZ+QTDPARTCES							
APGAR1: Asfixia moderada	-1,31	0,27	0,15	-8,53	<2e-16	0,024	
APGAR1: Asfixia leve	-2,93	0,05	0,13	-22,16	<2e-16		
APGAR1: Sem asfixia	-4,92	0,01	0,13	-38,23	<2e-16		
GRAVIDEZ: Única	-2,4	0,09	0,58	-4,13	3,56E-05		
GRAVIDEZ: Dupla	-1,52	0,22	0,6	-2,53	0,01		
QTDPARTCES: 4 a 6	0,78	2,18	0,71	1,09	0,28		
QTDPARTCES: 7 ou mais	-11,7	1,00E-05	663,2	-0,02	0,99		
QTDPARTCES: Nenhum	-0,02	0,98	0,11	-0,2	0,84		
IDANOMAL+QTDFILMORT+GRAVIDEZ							
IDANOMAL: Não apresenta	-3,57	0,03	0,13	-27,68	<2e-16	0,009	
QTDFILMORT: 2 filhos	0,49	1,63	0,24	2,02	0,04		

Tabela 10 continuação da página anterior

QTDFILMORT: 3 a 5 filhos	0,87	2,38	0,34	2,55	0,01	
QTDFILMORT: 6 ou mais filhos	-11,19	1,00E-05	707,8	-0,02	0,99	
QTDFILMORT: Nenhum	-0,07	0,93	0,13	-0,55	0,58	
GRAVIDEZ: Única	-2,98	0,05	0,58	-5,13	2,88E-07	
GRAVIDEZ: Dupla	-1,82	0,16	0,6	-3,01	2,59E-03	
IDANOMAL+QTDFILMORT+GRAVIDEZ						
IDANOMAL: Não apresenta	-3,59	0,03	0,13	-27,85	<2e-16	0,008
QTDFILMORT: 2 filhos	0,55	1,73	0,24	2,26	0,02	
QTDFILMORT: 3 a 5 filhos	0,98	2,68	0,34	2,9	0,004	
QTDFILMORT: 6 ou mais filhos	-12,15	0	1128	-0,01	0,99	
QTDFILMORT: Nenhum	-0,09	0,92	0,13	-0,66	0,51	
QTDPARTCES: 4 a 6	0,98	2,66	0,71	1,37	0,17	
QTDPARTCES: 7 ou mais	-12,09	1,00E-05	1294	-0,01	0,99	
QTDPARTCES: Nenhum	0,14	1,16	0,11	1,3	0,19	
IDANOMAL+GRAVIDEZ+QTDPARTCES						
IDANOMAL: Não apresenta	-3,59	0,03	0,13	-27,86	<2e-16	0,008
GRAVIDEZ: Única	-2,95	0,05	0,58	-5,09	3,53E-07	
GRAVIDEZ: Dupla	-1,75	0,17	0,6	-2,89	0,004	
QTDPARTCES: 4 a 6	1,02	2,76	0,71	1,42	0,15	
QTDPARTCES: 7 ou mais	-11,07	1,55E-05	760,8	-0,02	0,99	
QTDPARTCES: Nenhum	0,11	1,11	0,11	0,99	0,32	
QTDFILMORT+GRAVIDEZ+QTDPARTCES						
QTDFILMORT: 2 filhos	0,52	1,68	0,24	2,16	0,03	0,001
QTDFILMORT: 3 a 5 filhos	0,85	2,34	0,34	2,52	0,01	
QTDFILMORT: 6 ou mais filhos	-12,25	4,78E-06	1102	-0,01	0,99	
QTDFILMORT: Nenhum	-0,1	0,91	0,13	-0,76	0,45	
GRAVIDEZ: Única	-2,79	0,06	0,58	-4,82	1,43E-06	
GRAVIDEZ: Dupla	-1,54	0,21	0,6	-2,57	0,01	
QTDPARTCES: 4 a 6	0,64	1,9	0,71	0,9	0,37	
QTDPARTCES: 7 ou mais	-12,27	5,00E-06	1197	-0,01	0,99	
QTDPARTCES: Nenhum	0,15	1,16	0,11	1,37	0,17	

Tabela 11 – Coeficientes dos modelos estimados com 4 variáveis

Modelo	β	$EP(\beta)$	$\exp(\beta)$	z	$Pr(> z)$	R^2
APGAR1+IDANOMAL+QTDFILMORT+GRAVIDEZ						
APGAR1: Asfixia moderada	-1,13	0,32	0,16	-6,85	7,24E-12	0,027
APGAR1: Asfixia leve	-2,59	0,08	0,14	-18,20	<2e-16	
APGAR1: Sem asfixia	-4,53	0,01	0,14	-32,01	<2e-16	
IDANOMAL: Não apresenta	-2,57	0,08	0,14	-18,61	<2e-16	
QTDFILMORT: 2 filhos	0,1	1,11	0,25	0,41	0,68	
QTDFILMORT: 3 a 5 filhos	0,91	2,50	0,38	2,43	0,02	
QTDFILMORT: 6 ou mais filhos	-11,02	1,64E-05	680,90	-0,02	0,99	
QTDFILMORT: Nenhum	0,02	1,02	0,14	0,17	0,87	
GRAVIDEZ: Única	-2,59	0,08	0,58	-4,45	8,75E-06	
GRAVIDEZ: Dupla	-1,7	0,18	0,61	-2,81	0,005	
APGAR1+IDANOMAL+QTDFILMORT+QTDPARTCES						
APGAR1: Asfixia moderada	-1,11	0,33	0,16	-6,80	1,07E-11	0,026
APGAR1: Asfixia leve	-2,58	0,08	0,14	-18,15	<2e-16	
APGAR1: Sem asfixia	-4,55	0,01	0,14	-32,22	<2e-16	
IDANOMAL: Não apresenta	-2,59	0,08	0,14	-18,57	<2e-16	
QTDFILMORT: 2 filhos	0,14	1,15	0,25	0,56	0,57	
QTDFILMORT: 3 a 5 filhos	0,99	2,7	0,38	2,64	0,01	
QTDFILMORT: 6 ou mais filhos	-11,98	1,00E-05	1088	-0,01	0,99	
QTDFILMORT: Nenhum	-0,01	0,99	0,14	-0,08	0,94	
QTDPARTCES: 4 a 6	1,4	4,07	0,72	1,96	0,05	
QTDPARTCES: 7 ou mais	-12,98	2,00E-06	1570	-0,01	0,99	
QTDPARTCES: Nenhum	0,19	1,2	0,12	1,61	0,11	
APGAR1+IDANOMAL+GRAVIDEZ+QTDPARTCES						
APGAR1: Asfixia moderada	-1,12	0,33	0,16	-6,88	5,81E-12	0,027
APGAR1: Asfixia leve	-2,59	0,08	0,14	-18,42	<2e-16	
APGAR1: Sem asfixia	-4,52	0,01	0,14	-32,39	<2e-16	
IDANOMAL: Não apresenta	-2,62	0,07	0,14	-18,90	<2e-16	
GRAVIDEZ: Única	-2,57	0,08	0,58	-4,43	9,53E-06	
GRAVIDEZ: Dupla	-1,69	0,19	0,61	-2,78	0,01	
QTDPARTCES: 4 a 6	1,27	3,57	0,72	1,78	0,08	
QTDPARTCES: 7 ou mais	-12,19	5,11E-06	992,00	-0,01	0,99	
QTDPARTCES: Nenhum	0,15	1,17	0,11	1,35	0,18	
APGAR1+QTDFILMORT+GRAVIDEZ+QTDPARTCES						
APGAR1: Asfixia moderada	-1,31	0,27	0,15	-8,49	<2e-16	0,024
APGAR1: Asfixia leve	-2,93	0,05	0,13	-22,05	<2e-16	
APGAR1: Sem asfixia	-4,91	0,01	0,13	-37,97	<2e-16	
QTDFILMORT: 2 filhos	0,31	1,36	0,24	1,27	0,21	
QTDFILMORT: 3 a 5 filhos	0,97	2,65	0,37	2,60	0,01	
QTDFILMORT: 6 ou mais filhos	-11,73	8,06E-06	919,30	-0,01	0,99	
QTDFILMORT: Nenhum	-0,01	0,99	0,13	-0,07	0,95	
GRAVIDEZ: Única	-2,42	0,09	0,58	-4,17	3,02E-05	
GRAVIDEZ: Dupla	-1,57	0,21	0,60	-2,60	0,01	
QTDPARTCES: 4 a 6	0,79	2,21	0,71	1,11	0,27	
QTDPARTCES: 7 ou mais	-12,67	3,14E-06	1097,00	-0,01	0,99	
QTDPARTCES: Nenhum	0	1,00	0,11	0,01	0,99	
IDANOMAL+QTDFILMORT+GRAVIDEZ+QTDPARTCES						
IDANOMAL: Não apresenta	-3,57	0,03	0,13	-27,68	<2e-16	0,009
QTDFILMORT: 2 filhos	0,49	1,64	0,24	2,04	0,04	
QTDFILMORT: 3 a 5 filhos	0,88	2,41	0,34	2,58	0,01	
QTDFILMORT: 6 ou mais filhos	-12,17	5,19E-06	1169	-0,01	0,99	
QTDFILMORT: Nenhum	-0,08	0,92	0,13	-0,63	0,53	
GRAVIDEZ: Única	-2,99	0,05	0,58	-5,15	2,63E-07	
GRAVIDEZ: Dupla	-1,83	0,16	0,60	-3,04	0,002	
QTDPARTCES: 4 a 6	0,96	2,62	0,71	1,35	0,18	
QTDPARTCES: 7 ou mais	-12,13	5,40E-06	1277	-0,01	0,99	
QTDPARTCES: Nenhum	0,14	1,15	0,11	1,26	0,21	